AD-A277 407

AFOSR·TR· 94 0129   AUDITORY PATTERN MEMORY

Mechanisms of Temporal Pattern Discrimination by Human Observers

Robert D. Sorkin
Hearing Research Center
Department of Psychology
University of Florida
Gainesville, Florida 32611

15 February 1994

Final Technical Report for Period 1 October 1990 - 31 December 1993

AFOSR-91-0065

**DTIC
S ELECTE
MAR 2 8 1994
F D**

94 3 25 057

Prepared for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, DC 20332-6448

94-09339

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR· 94 0129

6a. NAME OF PERFORMING ORGANIZATION Hearing Research Center Department of Psychology	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NL
6c. ADDRESS (City, State and ZIP Code) University of Florida P.O. Box 112250 Gainesville, Florida 32611		7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-91-0005

8c. ADDRESS (City, State and ZIP Code) 110 Duncan Ave, Suite B115 Bolling AFB, DC 20332-0001	10. SOURCE OF FUNDING NOS.			
	PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2313	TASK NO. AS.	WORK UNIT NO.

11. TITLE (Include Security Classification)
Auditory Pattern Memory

12. PERSONAL AUTHOR(S)
Robert D. Sorkin

13a. TYPE OF REPORT Final Technical	13b. TIME COVERED FROM 90-10-1 TO 93-12-31	14. DATE OF REPORT (Yr., Mo., Day) 94-2-15	15. PAGE COUNT 51

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Auditory perception, temporal pattern discrimination group signal detection, signal detection with multi-element visual displays.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The project's first component studied how human listeners discriminate two temporal patterns of tones. The stimuli were non-speech, non-musical, tone sequences that conveyed arhythmic or partly rhythmic time patterns. The results indicated that the temporal pattern discrimination process depends on the timing of the sequences and whether they overlap in time. Listeners can perform the task very well when the patterns do not overlap and are presented to separate ears and/or at different tone frequencies. At long time separations, the listener's discrimination mechanism reduces the input information to two lists of intertone times, and the decision is based on the correlation between the lists (a temporal correlation process). At short time separations, the mechanism computes statistics based on the summed envelope of the two input waveforms (a single channel process). The project's second component consisted of analytical and computer modeling of multi-element detection systems. New results were obtained on the performance of statistically optimal detector arrays and on arrays that combine their binary outputs. The project's third component studied how observers process information in multi-element, visual signal detection. The results describe how performance depends on display element reliability, coding and arrangement.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr John F. Tangney	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5021	22c. OFFICE SYMBOL NL

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.   SECURITY CLASSIFICATION OF THIS PAGE

TABLE OF CONTENTS

1 OBJECTIVES AND STATUS OF THE RESEARCH EFFORT

1.1 Perception of temporal patterns

These experiments continued our studies of how human listeners discriminate temporal patterns (Sorkin, 1990; Sorkin and Montgomery, 1991). In the present experiments, we extended the pattern discrimination paradigm to cases when the tonal sequences were presented in different frequency or earphone channels and when the sequences overlapped in time. When the sequences were presented at separate times (or at precisely the same time), performance was very good. When the sequences overlapped in time, performance was very poor. These results are consistent with the operation of a discrimination mechanism (Sorkin, 1990) that has difficulty in resolving patterns that are presented at the same time.

In our experiments, the listener is presented with two arrhythmic tonal sequences. The series of time intervals between the tone onsets in each sequence, $<t_{1,1}, t_{1,2}, \ldots t_{1,n}>$, and $<t_{2,1}, t_{2,2}, \ldots t_{2,n}>$, define the temporal patterns to be discriminated. On half of the trials (SAME trials) these two temporal patterns are linearly related and hence perfectly correlated ($t_{1,j} = at_{2,j} + b$ for all j), and on half of the trials (DIFFERENT trials) the patterns are not perfectly correlated. The listener must report whether the temporal patterns were the same or different (see figure 1). An important experimental variable is the correlation between the sequence patterns on DIFFERENT trials, ρ_{diff}. The task is easiest when ρ_{diff} equals 0 and increases in difficulty as ρ_{diff} approaches one.

A number of factors affect temporal pattern discrimination in addition to the temporal correlation. These include the number and spectral properties of the marker tones, the temporal properties of the patterns, and the location of the information within the patterns (for recent studies, see Espinoza-Varas and Watson, 1986; Bregman, 1990; Hirsh, et al., 1990; Kidd and Watson, 1992; Monahan and Hirsh, 1990; Schulze, 1989; and Sorkin, 1990). The time interval between pattern onsets is a potentially important factor in affecting pattern processing, but it has not received much experimental attention, particularly at very brief intervals.

Sorkin (1990) proposed the temporal pattern correlation (TC) model for describing how listeners discriminate between monaural temporal patterns. According to this model, listeners discriminate between arrhythmic tonal sequences by estimating the correlation between the temporal patterns formed by the two sequences. The basic stimulus is assumed to be the series of times between the onsets of the tones in each sequence: The listener extracts and stores two lists of interonset times, one for each sequence, and then estimates the correlation, $\rho_{t1, t2}$, between the two lists. The TC model assumes that the system discards information about the stimulus sequences that is irrelevant to the correlation computation, such as overall changes in the presentation rate or

4

·the frequency of the tones.

The TC model allows one to predict the effect of transform-
ing or distorting the time intervals in each pattern. For exam-
ple, if all the interonset times in one of the sequences were
multiplied by a constant (thereby producing a uniform temporal
expansion of that sequence), the correlation computed between the
sequences on a trial would be unchanged. Listeners employing a
TC mechanism should be relatively insensitive to such a manipula-
tion. Similarly, adding or subtracting a constant time to all
the intervals in one of the sequences should have little effect
on the correlation calculation and hence on discrimination per-
formance. The effect of the latter manipulation would depend on
the level of internal noise in the TC system.

==

(a) SAME sequences   ·(b) DIFFERENT sequences

sequence 1   sequence 1

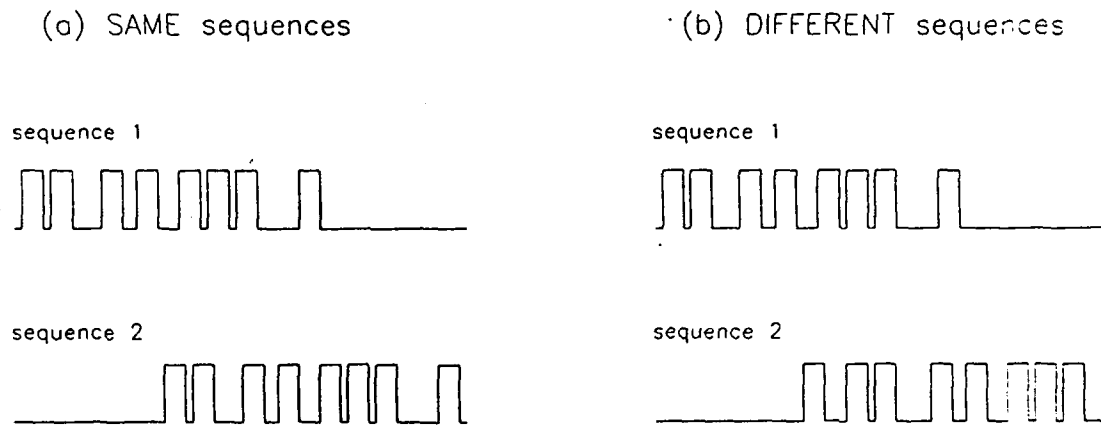sequence 2   sequence 2

Figure 1. Envelope gating functions for typical tone sequences on SAME (a) and DIFFERENT (b) trials.

5

Sorkin and Montgomery (1991) showed that listeners could perform the discrimination task at a level that was well above chance, when uniform time transformations were made to one of the two patterns. All tones in their experiments were 1000 Hz; the sequences were presented monaurally and the time separation between the end of the first sequence and the beginning of the second sequence was approximately 800-ms. Performance decreased when the second sequence was compressed or expanded in time, and depended on the magnitude of the time transformation between the two sequences. The size of the decrease in performance ranged from 0 to 2 d' units over transformations of 0.6 to 1.6. The results supported the assumption that there was an internal noise proportional to the absolute magnitude of the transformation difference.

These types of temporal transformations are common in speech and music perception. Evidence supporting the model and the relationship between temporal pattern discrimination and speech recognition also has been obtained with hearing impaired listeners using cochlea prostheses (Collins and Wakefield, 1992). Collins and Wakefield found that their observer's ability to discriminate temporal patterns depended on the temporal correlation between the two sequences, as predicted by the TC model. In addition, they reported that the observers' ability to discriminate arrhythmic sequences was positively correlated with the observers' speech recognition performance.

1.1.1 Discrimination of delayed temporal sequences (Sorkin, Montgomery, and Sadralodabai).

In experiment 1, we asked listeners to perform the temporal pattern discrimination task when the tonal sequences were presented monaurally and dichotically in different frequency bands. We wished to test how temporal pattern discrimination performance depended on the intersequence delay between the patterns. Recall that the TC mechanism extracts and stores a list of interonset times for each sequence, and then estimates the correlation between the two lists. Assuming that the time extraction process can be performed when the patterns are presented in two separate frequency and/or earphone channels, listeners should be able to perform the two-channel pattern discrimination task under different intersequence delay conditions. We were particularly interested in the performance that would result when the sequences overlapped in time. In experiment 2, we imposed a small, random temporal transformation on the second of each pair of tonal sequences. The operation of the assumed two-channel TC mechanism should be insensitive to such small random expansions or compressions, whether or not the sequences are presented in separate channels or overlap in time.

The subjects were undergraduate students at the University of Florida. They were paid an hourly wage plus an incentive for correct responses. The subjects had normal hearing and performed the tasks for approximately 2 h per day, 3 days per week. Subjects were seated in a double-walled acoustically insulated

6

chamber. The stimuli were presented monaurally or dichotically via TDH-39 headphones. The conditions were tested in blocks of 100 trials; 8 to 12 blocks were completed in a session. Except in the uncertain time transformation conditions, all independent variables were held constant within a block of trials. Feedback about the correct response was provided after each trial.

The subjects compared pairs of tone sequences composed of 9 sinusoidal bursts of 30-ms. After listening to each pair of sequences, the subject had to indicate whether or not the temporal pattern of intertone time intervals was the same or different for the two sequences. On a random half of the experimental trials, the temporal patterns were the same, p_{same} = 1.0, and on the remaining trials the patterns were different, p_{diff} = 0.2, in a block of 100 trials. The average duration of a sequence of tones was 670-ms. The mean intertone interval (time between tone onsets) was either 80, 120, or 160-ms, and the standard deviation was 30-ms; the minimum tone interonset time was 32-ms. The mean and standard deviation of the intertone intervals and the correlation between temporal sequences were controlled by a process described in Sorkin (1990). There were two groups of subjects in experiment 1. Group 1 was composed of two male and two female undergraduates. Subjects in this group ran all conditions at a mean intertone interval of 80-ms. The dichotic conditions in experiment 1 were repeated at a later date with a second group of subjects (Group II, composed of one male and one female undergraduate). Subjects in Group II were tested at mean intertone intervals of 80, 120, and 160-ms.

In order to compare discrimination when the patterns overlapped in time, the sequences were presented at different frequencies and to different earphone channels. The tone bursts in the first sequence were 1000 Hz and the tones in the second sequence were 2300 Hz; they were set approximately equal in loudness at 71 dBA and 68 dBA, respectively. In the monaural conditions, both sequences were presented to the right ear. In the dichotic conditions, the first sequence was always presented to the right headphone and the second to the left. The onset of the second sequence (i.e., the first tone) was presented at intersequence intervals (ISIs) of from 0 to 2.5 seconds after the onset of the (first tone in the) first sequence. All tone bursts were shaped by a 4-ms linear rise and decay.

Experiment 1. Effect of two-channel presentation

The purpose of this experiment was to examine how pattern discrimination depended on the intersequence interval between the two sequence starting times. In addition, we wished to extend the pattern discrimination task to the case when the sequences were presented in different frequency bands and in different earphone channels.

In the monaural condition, the second sequence (2300 Hz tones) began in the right earphone channel at a fixed time (intersequence interval) after the onset of the first sequence

(1000 Hz tones) in the right channel. The time intervals were 0,
10, 20, 100, 350, 900, and 2500-ms. For the dichotic condition,
the second sequence (2300 Hz) was in the left earphone channel.

Figure 2 shows the effect of intersequence interval on the
average performance of four listeners. The circle symbols (solid
lines) show performance in the monaural conditions and the square
symbols (dashed lines) show the dichotic conditions. The verti-
cal bars are the average of the standard errors of the mean for
the listeners. There were no systematic differences between the
dichotic and monaural conditions. Best performance was obtained
at an intersequence interval of 0 ms., when the two patterns
completely overlapped. The data from all of the listeners showed
that performance began to deteriorate at an intersequence inter-
val of 20-ms and poorest performance was obtained between approx-
imately 100 and 400-ms. At these intervals, the sequences over-
lapped (on average) 85% and 40%, respectively. Performance
improved when the delay was increased to 900-ms, then leveled off
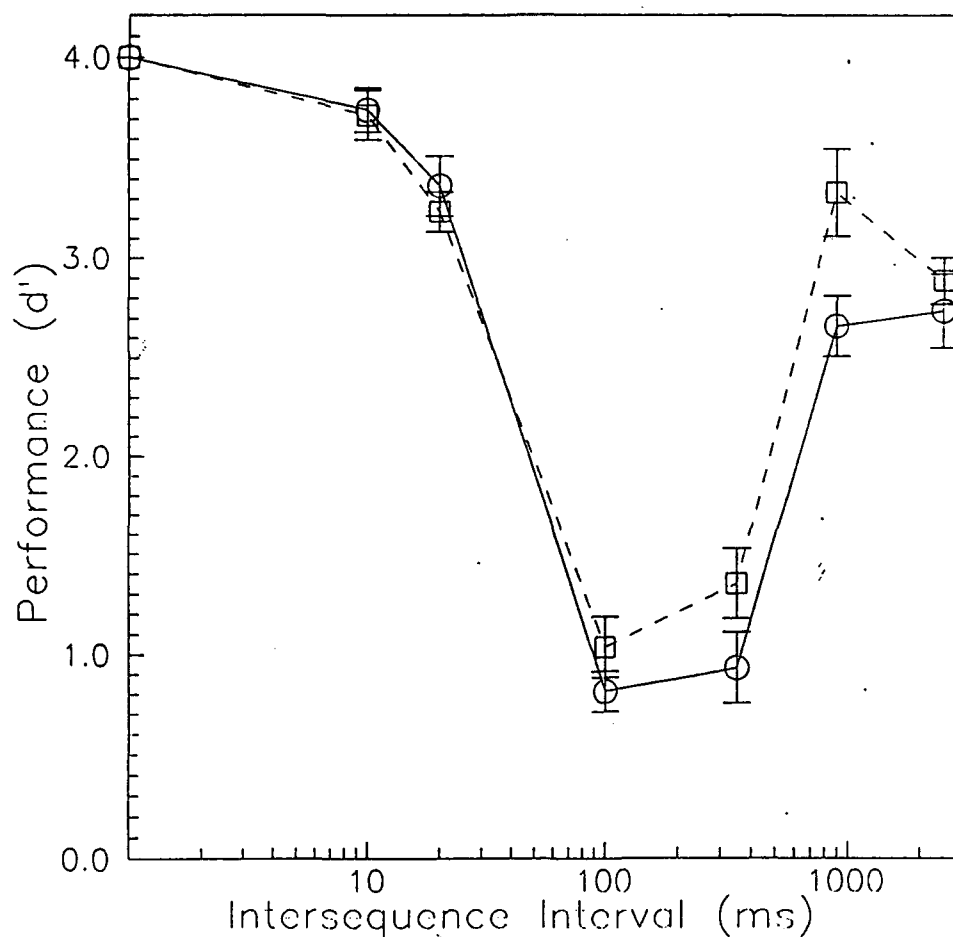or decreased at a delay of 2500-ms.
===

Figure 2. The average performance (d') of the listeners in Group I of experiment 1, as a function of the interse-
quence interval. The circle symbols are the data from the monaural conditions and the square symbols are the
data from the dichotic conditions. The brackets show the average standard error of the mean for the four lis-
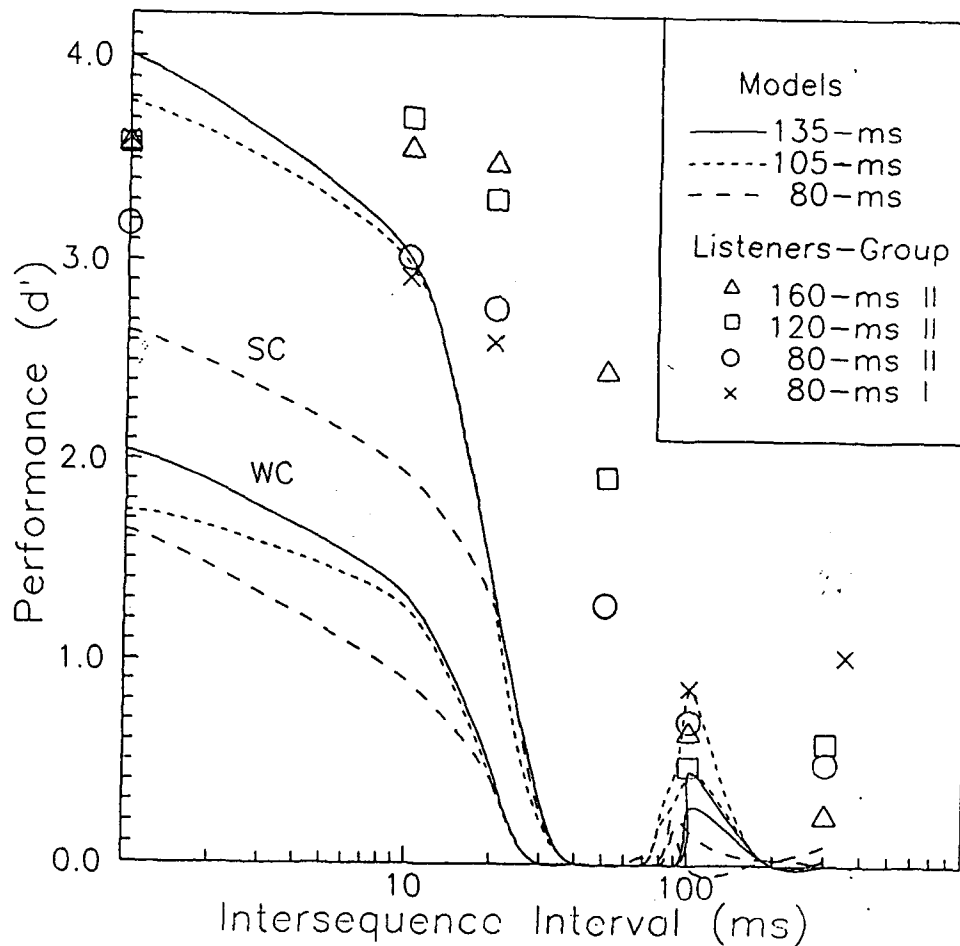teners. The value of the intersequence interval at the origin of the graph is 0-ms.

8

Figure 3. The average performance (d') of the listeners in Groups I and II in the dichotic conditions of experiment 1, plotted as a function of the intersequence interval. The data points (symbols) show listener performance at mean intertone intervals of 80, 120, and 160-ms. The solid and dashed curves show the performance of two hypothetical discrimination mechanisms: the Single-Channel (SC) mechanism (upper curves) and the Waveform-Correlator (WC) mechanism (lower curves) evaluated at mean intertone intervals of 80, 105, and 135-ms (see text).

We ran two additional subjects (Group II) at different intertone intervals, to check whether the performance drop at short intersequence delays was specific to the particular inter-

tone intervals used. The points plotted in figure 3 show the average performance of the subjects in Group II. The circle, square and triangle symbols show, respectively, performance at intertone intervals of 80, 120, and 160-ms for the dichotic conditions of experiment 1. The x-symbols show the average performance of the Group I subjects at 80-ms in the dichotic conditions. The standard errors are left off for clarity. The 80-ms data for the two groups are consistent. The data at 120-ms and 160-ms show higher levels of performance, but there are a few reversals. For all conditions, performance was lowest at inter-sequence intervals of 100-ms and 300-ms. We also ran some other combinations of intertone interval, mean tone duration, and standard deviations of the intertone interval of values other than 30-ms (not plotted). Increasing the average duration of the intertone interval (by increasing the intertone gaps or the tone durations) resulted in improved performance at intermediate intersequence delays, but performance always dropped to a minimum level by an intersequence interval of 300-ms.

Performance at the 900-ms (and longer) conditions replicated our earlier results, which we have interpreted as supporting a TC mechanism. However, the TC model contains no assumptions about the effects of intersequence delay (or pattern overlap) and so does not predict the poor discrimination performance at delays between approximately 20 and 400-ms. This poor performance may be due to an inability to extract the information from each channel needed to compute a temporal correlation. If the TC mechanism cannot function when there is pattern overlap, we need to explain how the task is performed when the patterns overlap almost perfectly, i.e., when the intersequence delay was <u>less</u> than 30-ms and performance was very good. We shall consider two possible mechanisms for accomplishing this function.

Candidate Discrimination Mechanisms

One possible mechanism for performing the discrimination task at short delays and when the signals overlap in time, is a simple, single-channel mechanism. The single-channel mechanism could be supplanted by the TC process, when the delays were long enough to allow the system to separately process the inputs on the two channels. We assume a very simple mechanism: the single-channel mechanism sums the signals in the two channels prior to the extraction of any intertone timing information. This yields a combined signal that is the sum of the envelope gating func-tions of the two sequences. Information about whether the two sequences are the same or different is obtained from statistics based on the properties of the summed single-channel signal.

Consider some properties of the summed envelope gating function on SAME and DIFFERENT trials, when the intersequence delay is zero. On SAME trials, the two channels contain the same gating pattern, so the resulting summed signal would consist of 9 tones of 30-ms duration having a mean intertone <u>gap</u> of 50-ms and a mean intertone standard deviation of 30-ms. On DIFFERENT trials, however, the summing operation would produce a signal

10

with more than 9 tones and with a mean tone-on duration greater than 30-ms. These statistics, the number of discrete tones and the mean tone-on (or off) duration, could provide information about the likelihood that the two sequences had been generated by DIFFERENT or SAME trials. As the intersequence interval was increased, the effectiveness of the statistics would decrease.

A computer simulation of such a single-channel (SC) mechanism was implemented. On every trial, the two statistics describing the summed signal were computed and combined. The discrimination performance of the SC model is shown as the three upper curves in figure 4, for mean intertone intervals of 80, 105, and 135-ms. The important parameter of the model is the assumed jitter in the system's estimate of the onsets (and offsets) of the resulting gating function. For all the curves shown, the standard deviation of this jitter was set at 4-ms. The model's performance dropped rapidly after an intersequence delay of 10-ms, for all three values of mean intertone interval. Larger jitter values resulted in greatly decreased performance at all intersequence delays.

The three lower curves on figure 4 show the performance of another simple mechanism: a simple waveform correlator (WC). We assume that this mechanism can obtain the temporal gating functions from the two channels, multiply the two functions together, and then integrate the resulting waveform over the duration of the patterns. The jitter in the system's estimate of the onsets (and offsets) of the separate channel gating functions was set at 4-ms. As in the SC case, this is the major parameter of the model. From figure 4, the performance of this mechanism is poorer than that of the SC mechanism. For both models, there were small increases in performance at intersequence intervals that were approximately equal to the period of one intersequence delay (when the second tone in one pattern was in rough alignment with the first tone in the other pattern). Otherwise, model performance fell to a low or chance value by the time the intersequence interval reached approximately 30-ms.

In order to provide further comparisons of these models with the performance of human listeners, in experiment 2 the temporal properties of the sequences were randomly varied over trials. We expected these manipulations to produce large effects on the performance of the candidate models, but we were not sure what effect the manipulations would have on the performance of human listeners at short intersequence intervals.

Experiment 2. Interaction of intersequence delay and temporal transformation

The sequence discrimination task in experiment 1 was modified in a manner designed to differentially affect the operation of the hypothetical mechanisms. This manipulation was a uniform temporal compression or expansion of all of the times (marker tones and gaps) in the second sequence, similar to that described in Sorkin and Montgomery (1991). The magnitude of the transfor-

11

mation was fixed within a trial but varied randomly over the
trials within a block.

The temporal correlation mechanism (i.e., a two-channel
process) is relatively insensitive to this manipulation (Sorkin
and Montgomery, 1991). On the other hand, the simple WC and SC
mechanisms should be highly sensitive to this manipulation be-
cause of their dependence on the temporal coherence of the two
patterns on SAME trials.

Experiment 2 was similar to Experiment 1, except for the
additional expansion/compression manipulation. Performance was
assessed at the same p_{diff} and intersequence intervals as in
Experiment 2. All marker tone durations and intertone gaps in
the second sequence of tones were multiplied by a factor chosen
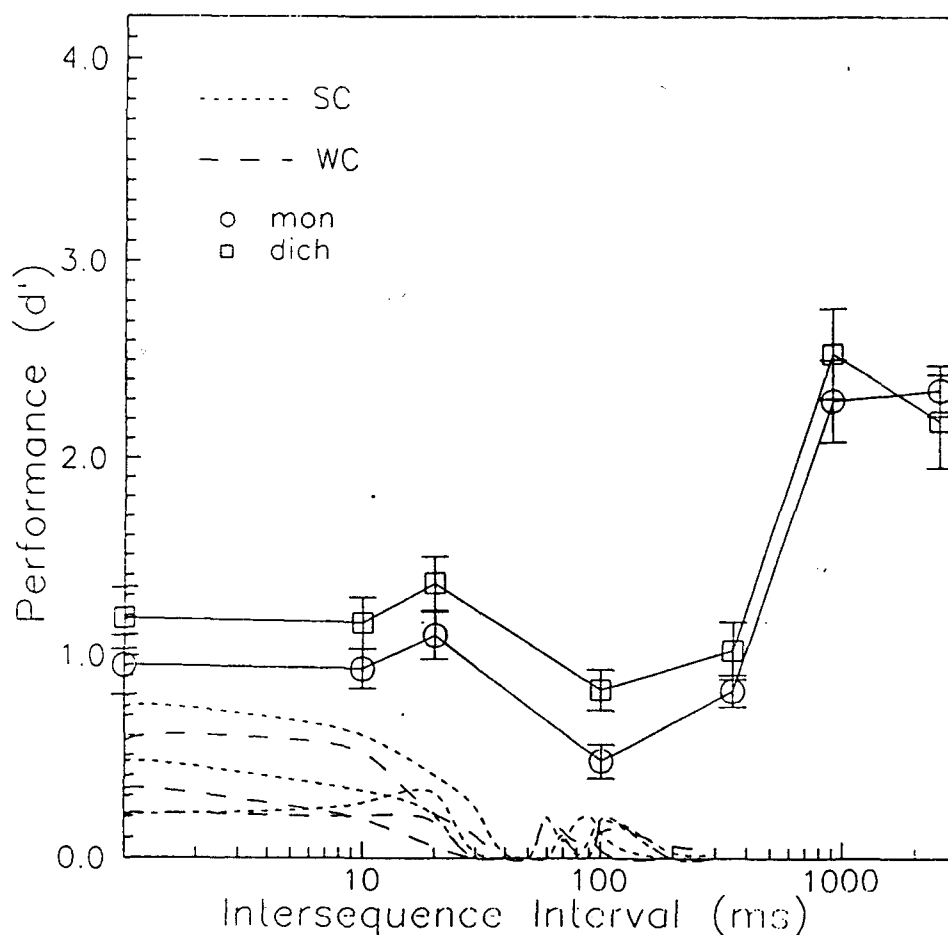from among the values: 0.8, 0.9, 1.0, 1.1 or 1.2. On each trial

Figure 4. The average performance (d') of the listeners in Group I, experiment 2 (random temporal transforma-
tion) and of two hypothetical discrimination mechanisms, as a function of the intersequence interval. The circle
symbols are the human data in the monaural conditions and the square symbols are the human data for the
dichotic conditions; the mean intertone interval was equal to 80-ms. The brackets show the average standard
error of the mean for the four listeners. The dashed curves show the performance of a hypothetical Single-
Channel (SC) mechanism (smaller dashes) and a Waveform-Correlator (WC) mechanism (larger dashes),
evaluated at a mean intertone interval of 80, 105, and 135-ms.

12

of the experiment, this factor was chosen randomly from the list of 5 values, and uniformly applied to all the time intervals within the second sequence. The probability of a particular transformation being chosen was 0.2. The subject was required to indicate whether the temporal pattern of tones was the same or different, ignoring whether the overall tempo of one pattern had been scaled faster or slower by the time transformation.

Figure 4 shows the effects of the expansion manipulation and intersequence interval on the average performance of the four subjects for the monaural and dichotic conditions. The circle symbols (solid lines) show performance in the monaural conditions and the square symbols (dashed lines) show performance in the dichotic conditions. Performance at intersequence intervals of from 0 to 350-ms was poor and relatively uniform over intersequence interval; poorest performance was at 100-ms. All listeners had their highest performance at an intersequence interval of 900-ms. It is evident that the random time transformation led to poorer performance than the no-transformation condition. The performance drop produced by the time transformation at long intersequence intervals was very small, consistent with our previous results and with the predictions of the TC model (Sorkin and Montgomery, 1991). On the other hand, the time manipulation resulted in drastically reduced performance even at very small intervals.

When the performance of the SC and WC mechanisms was simulated in the random time transformation, very poor performance resulted. The simulation results are shown as the solid and dashed lines on figure 4. The poor performance of these mechanisms is consistent with our expectations about their sensitivity to manipulations that disturb the temporal coherence of the sequences on SAME trials.

Conclusions

The results of experiment 1 indicated that listeners could discriminate between two temporal patterns, even when the two patterns were defined by (iso-frequency) tone sequences presented at different frequencies and to different ears. Presenting the sequences dichotically did not have much effect on performance. The good performance observed at long intersequence intervals under both the no-transform and random-compression/expansion manipulations, was consistent with the operation of a TC mechanism. Listener performance was very poor at intermediate intersequence intervals when the sequences overlapped. We concluded that the TC mechanism does not operate when the patterns overlap.

Listener performance was very good at very short intersequence intervals (when the two sequences were almost coincident), so long as the random compression/expansion manipulation was not applied. Therefore, it is necessary to postulate a mechanism that (a) can discriminate the sequence patterns at very short (but not at long) intersequence intervals, and (b) is sensitive to random compression and expansion of one of the patterns. The

performance of two candidate mechanisms, the single-channel and the waveform correlation mechanism, were evaluated and compared to the observed data. The performance of these simple mechanisms was qualitatively similar to the listeners' behavior. Both mechanisms performed poorly when the intersequence interval was longer than about 30-ms, and both performed poorly when a random transformation was imposed.

These results suggest a two-part mechanism: When the time interval between sequence onsets is long enough that the patterns do not overlap, a likely mechanism is temporal pattern correlation. Under these conditions, the important stimulus information is conveyed by the pattern of times between the tone onsets in each sequence. However, when the time interval between sequence onsets is brief, the pattern discrimination process is likely based on the combined (summed or multiplied) inputs from the two input channels.

Why can't the temporal pattern correlator function when the sequences overlap? Our argument goes to the basic nature of the TC mechanism. The hypothesized TC mechanism exemplifies a type of stimulus processing that Durlach and Braida (1969) have called "context coding". This type of processing requires the system to abstract data from each stimulus sequence (i.e., the tone inter-onset times), and then store this encoded data (i.e., as an ordered list of times) prior to performing the correlation operation. The encoded information does not require large capacity storage, and may be available in memory for many seconds prior to processing and decision (Durlach and Braida, 1969; Sorkin, 1987). We believe that the attentional demands imposed by stimulus processing and encoding, limit system operation to a single-channel mode. As a result, in order for the temporal pattern correlation mechanism to function effectively, the stimuli have to be presented sequentially in time. We suspect that this may be a general requirement for processing signals in the context-coding mode.

1.1.2 Effects of rhythmicity on temporal pattern discrimination (Sorkin and Sadralodabai).

The rhythmic aspect of a stimulus is an important property of a temporal pattern. We have begun to analyze the effect of rhythmic properties on pattern discrimination, in the context of the TC model. Recently, we reported (Sadralodabai and Sorkin, 1992) on a preliminary study of the effect of rhythmicity on the discrimination of temporal patterns. Observers were presented with two sequences of 12 tones and asked to discriminate whether the two patterns were the same or different. The duration and the frequency of tones were 25 ms and 1000 Hz respectively. As in our other experiments, the temporal pattern of each sequence was determined by the intertone time intervals.

Two kinds of correlation were important in this experiment: One was the sequence correlation, ρ_{ex}, the correlation between the two 12-tone temporal patterns, as defined earlier. The second

type of correlation, the rhythmic correlation, p_{rhy}, was defined as the correlation between the temporal patterns of successive 4-interval subsequences within the 12-tone sequence. We used p_{rhy} as a measure of the rhythmicity of the sequences. For example, a rhythmic correlation of 0 indicates no repetition of sub-patterns within a given sequence, and a correlation of 1 indicates complete repetition of the sub-patterns within a sequence.

The control condition in this experiment replicated the original correlation experiments, i.e., p_{rhy}=0, or no repetition within the sequences. Values of the sequence correlation were 0, 0.4, and 0.8. The mean and standard deviation of the intertone time intervals were 50 and 35 ms respectively. Performance (d') decreased as the sequence correlation increased, consistent with the earlier results. The TC model was fitted to this data and the internal noise was estimated for each listener based on their performance. Estimated values of σ_{in} for observers 1, 2, and 3, respectively, were 19-ms, 22-ms, and 19-ms.

We then tested performance in the experimental condition, with a rhythmic correlation, p_{rhy}=1. That is, there were 3 repetitions of the 4-interval subsequences within the sequence (the last repetition contained only three intervals). The sequence correlation was varied from 0 to .8, in steps of 0.2. As can be seen by the plotted points in figure 4, performance was very good and decreased as the sequence correlation increased.

We constructed a simple extension of the TC model to this task, using the following argument: Normally, there are two lists of 11 intertone times that may be used to estimate the correlation between the temporal patterns. When there are repeating patterns within the sequence, there will be fewer (independent) intervals are available for the correlation estimate. In the p_{rhy}=1 case, there are only 4 independent intertone time intervals, although this pattern of four intervals repeats 3 times. Thus, when the listener estimates the correlation in the p_{rhy}=1 case, only 4 intertone times may be used instead of 11. This results in an increase in the variance of the estimate of the between-sequence correlation, and hence a potential decrease in performance. However, repeating the patterns within a sequence yields a reduction in the effect of the observer's internal noise, because the observer's estimates of the 4 intertone times within a repetition becomes (statistically) more reliable. Thus, according to the simple extension of the TC model: in the repetition condition the effective n is 4, rather than 11, and the effective internal noise (σ^2_{in}) is 1/3 of what it was in the non-repetition condition.

The model's predictions are shown as the smooth curves in figure 4. The improvement in performance due to the rhythmicity of the sequences was much better than predicted by the simple TC model. We also examined performance at rhythmic correlation values of 0, .5, 1, and at sequence correlations of 0 and .4. Most of the improvement in performance seemed to occur when the

rhythmic correlation was greater than 0.5. Results at an inter-
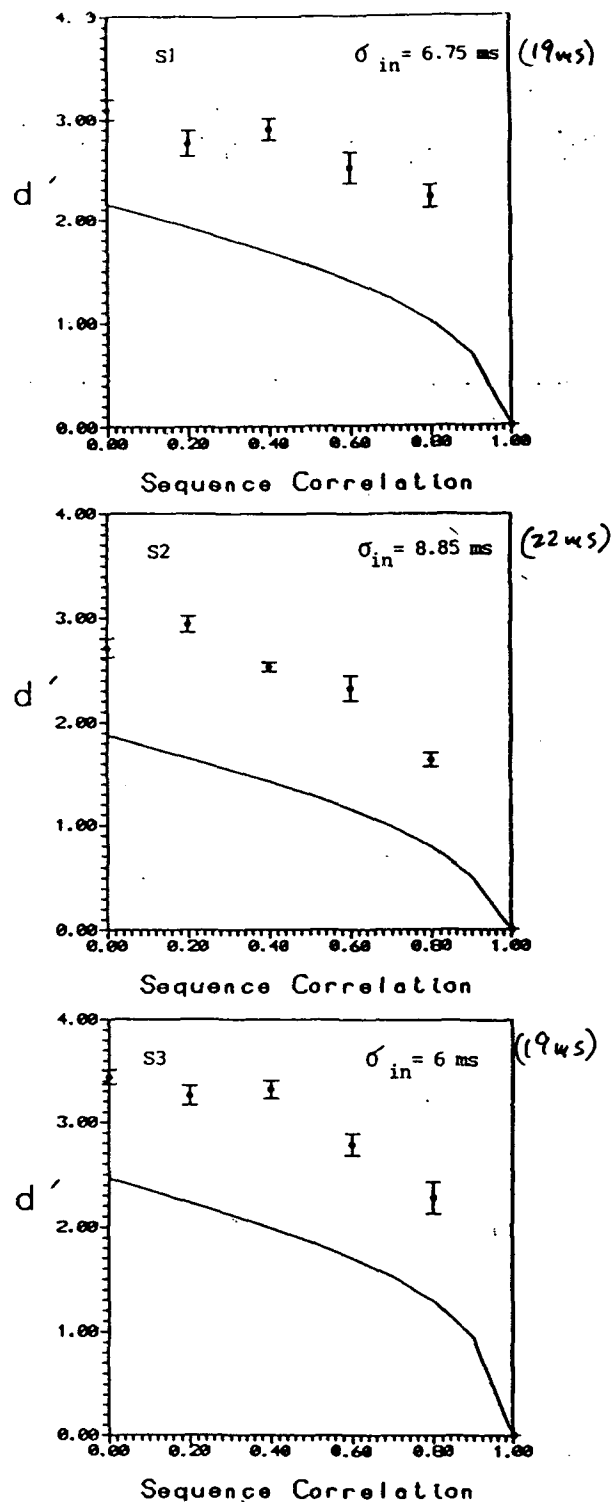tone interval of 100-ms also were consistent with these results.

Figure 5. The performance of three observers in the $p_{rhy}=1$ condition (circle symbols). The brackets show plus and minus one standard error of the mean. The smooth curve is the performance of the observer based on the TC model using a value for the observer's internal noise that was estimated from performance in a separate $p_{rhy}=0$ condition.

From these experiments, we conclude that the presence of

16

rhythmicity plays an important role in a listener's ability to discriminate between two temporal patterns. Further experiments will attempt to revise the model so that it can capture the effects of rhythmic properties of the patterns. It appears that (when $p_{rhy}=0$) the observer may be using a non-optimum strategy for deciding if the sequences are different; that strategy results in an improvement in performance when there is information that reduces the size of the ensemble of possible sequences (e.g. when $p_{rhy}>0$). One possibility is to construct conditions for which $p_{t1,t2}$ is not an optimum strategy and in which the observer may use information about the possible sequences on a trial.

We have begun a series of experiments to directly assess the effect of important task variables on the discrimination of rhythmicity. We continue with our assumption that the rhythmicity of a pattern is related to the correlation between temporal units within the pattern (as defined by p_{rhy} in a pattern that has partially repetitive cycles of m subpatterns of size k, with a uniform correlation between cycles). The observer's task in our experiments, will be to discriminate which of two patterns is more rhythmic. Our initial experiments indicate that observer's have no trouble with this two-interval-forced-choice task, and that adaptive techniques provide reliable estimates of performance.

1.1.3 Effect of temporal position and temporal context (Sorkin and Sadralodabai).

One weakness of the temporal correlation model is that it ignores effects occurring at different temporal or serial positions within the serial pattern. This insensitivity to time-position is a consequence of the assumption that pattern information is encoded independently of the location of that information within each sequence. Previous reviewers of our research have pointed out that this lack of sensitivity to conditional time-position properties may not be plausible, given what is known about speech and musical perception. For example, two patterns that have large and distinctive gaps near the end, and relatively uncorrelated patterns throughout the rest of their patterns, will probably be judged more similar than two patterns that have more uniform distributions of gaps, but a higher statistical similarity, i.e., temporal correlation (see Devenyi and Hirsh, 1975; Espinoza-Varas and Watson, 1986; Hirsh et al. 1990; and Watson et al. 1975, 1990). It is evident that a model that relies on a temporal correlation parameter that is uniformly defined over the pattern duration, probably will not be able to adequately specify the discriminability of the patterns.

To remedy this weakness of the TC model, we have begun to directly study the distribution of an observer's responses (different/same) as a function of both the position and the properties of the temporal intervals in the two stimulus sequences (Sadralodabai and Sorkin, 1993). This analysis is similar to those by Berg (1989, 1990), Berg and Green (1990), Lutfi (1989, 1990, 1992), and Sorkin et al. (1987), using the sample-discrimi-

17

nation procedure. Although our procedure is not formally identical to the sample-discrimination procedure, these techniques will enable us to determine the differential weight employed by observers at different positions in the sequences.

On each trial, the observer's response and the sequence of intertone intervals in each sequence, is recorded. We compute a COSS-type function of the difference (and the product) of the corresponding intervals in each sequence. Specifically, we compute the probability that the observer has responded 'different', conditional on the magnitude of the difference between the intertone intervals at that serial position, and conditional on the magnitude of the product of the intertone intervals at that serial position.

That is, for DIFFERENT trials, and across all values of $|t_{2,j}-t_{1,j}|$ for $j \neq i$, we will compute (for each position, i):

$$p(\text{respond "DIFFERENT"} \mid |t_{2,i}-t_{1,i}|) \tag{1}$$

and

$$p(\text{respond "DIFFERENT"} \mid t_{2,i} \cdot t_{1,i}) \tag{2}$$

We assume that the observer's decision on each trial is based on either

$$\Sigma a_i(|t_{2,i}-t_{1,i}|) \quad \text{or} \quad \Sigma a_i(t_{2,i} \cdot t_{1,i}). \quad \text{(The latter statistic}$$

is a version of the TC model.) We use the standard deviation of the resulting distributions as an **estimate** of the observer's decision weight at position i. (The reader may wonder whether the properties of the resulting distributions can be used to determine **which** statistic was being used by the observer. From simulations, we know that the standard deviation of the difference and product distributions is approximately the same. Although the shape of the distributions are different, the number of trials required to tell which statistic was used probably is not feasible in a human experiment.)

Figure 6 shows some data obtained on a group of listeners using sequences of 4 and 8 tones, and analyzed with the modified COSS procedure. Relative weight is plotted as a function of the serial position of the interval in the sequence. The data indicate that the first position had the greatest influence on the listener's response. We have begun to study the changes in the serial position weights as a function of having a cycle of four intervals repeat within the sequence. The results so far indicate that the the weighting pattern depends on the rhythmicity of the sequence as well as on the mean duration of each interval.

We plan to perform these analyses using sequences in which the intertone intervals have non-uniform means or non-uniform standard deviations, at different serial positions in the sequence. Sequences generated by the latter procedure will have
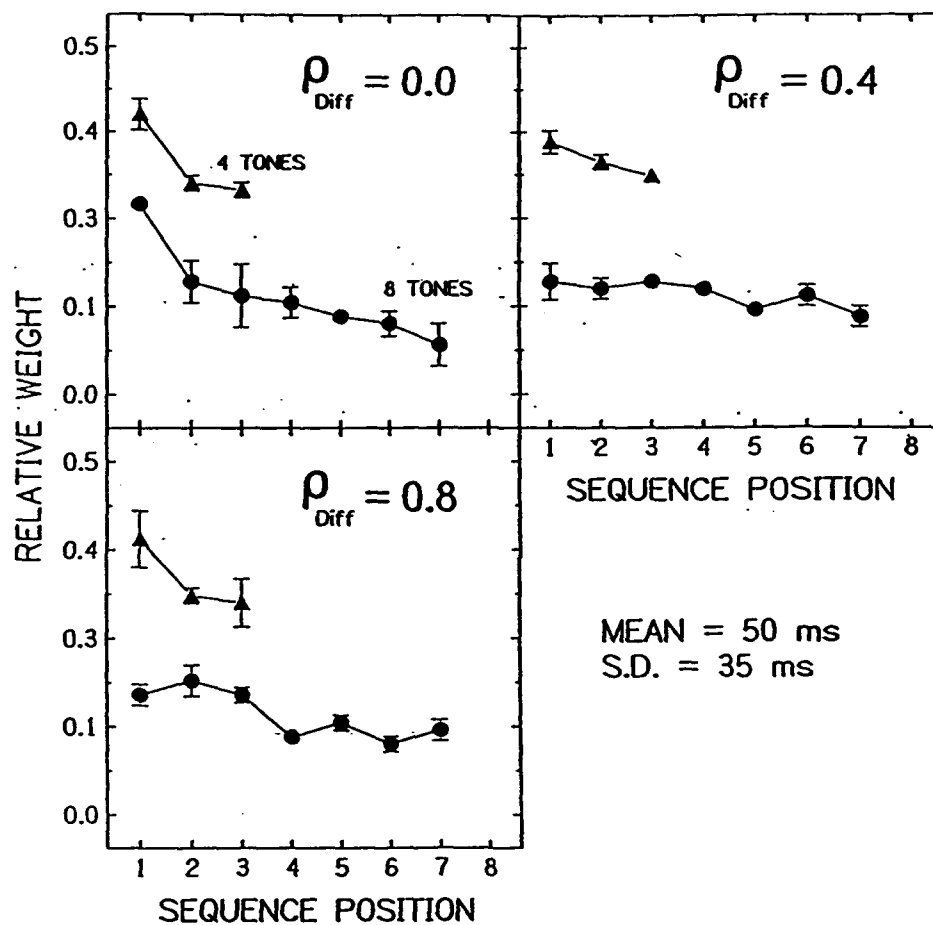
Figure 6. Average listener decision weights as a function of interval (serial) position for 4 and 8-tone sequences, for three different values of rho-different, and a mean gap of 50-ms and gap std. dev. of 35-ms.

serial positions that contain more information relevant to the task (in the sense of Lutfi's 1992 analysis and his Proportion of Total Variance hypothesis). Such positions should show higher observer weights than less variable intervals. We can also test whether the _distinctiveness_ of the interval in the sequence, rather than its informativeness or serial position, commands higher observer attention. Sequences will be constructed in which the intervals in some serial positions have higher mean durations; these positions should show higher observer weights than the positions having shorter mean intervals, in the sense of Kidd and Watson's (1992) Proportion of Total Duration hypothesis. These experiments should provide specific, quantitative data on the effects of serial-context factors on temporal pattern discrimination.

Finally, the temporal pattern correlation model has not been tested with specific subsets of temporal patterns. For example, Povell and Essens (1985) and others have argued that there is a natural organization or structure to certain temporal sequences, depending on the relationship between the position of occurrence of the tones in the sequence and the basic sequence timing. Suppose that the duration of the base cycle of a repeating sequence is 760-ms, each containing 8 tones of 40-ms duration, and the smallest inter-tone gap is 40-ms. Any tone must start on one of the 16 possible starting times defined by those 40-ms (discrete) periods. Assume that all patterns have a tone at the first period. Certain sequences, by virtue of the specific starting times of the tones, will be perceptually more 'structured' than others. We will conduct pattern discrimination experiments with different subsets of these fully random sequences, using different algorithms for selecting the patterns, such as for metricity and nonmetricity. Using the Pattern Correlation model, we will evaluate the statistical and empirical aspects of these effects.

1.2   Analysis of Group Detection Systems

We have been using the theory of signal detectability to develop models for describing how groups of detection systems can detect signals. These models are based on the theory of signal detectability, specifically on multi-channel auditory detection (Berg, 1990; Green, 1992; Durlach et al., 1986). The models enable us to make quantitative predictions relating group signal detection performance (accuracy, d'_{group}; bias, β_{group}; and efficiency, η_{group}) to a group's size, the mean and variance of member d', the correlation among member judgments, the relative influence of members on the decision), the group decision rule, and the degree of member interaction. These analysis are relevant to the training of groups, teams, and crews, as well as to the design of systems composed of human operators and machine detectors, such as alarm and alerting subsystems.

1.2.1 Analysis of the Ideal Group (Sorkin and Dai, in press).

A simplified concept of the multi-channel detection/decision process is illustrated by the system shown in figure 7. This system is composed of a group of detectors which must decide whether a signal or no-signal event was present on a trial. Each detector monitors a distinct channel and each channel is subjected to several noise sources: One of these sources is unique to each detector (in the figure: n_1, n_2, n_3), and the other sources are common to two or more detectors (e.g. $n_{1,2,3}$, $n_{1,3}$). Each detector computes a statistic, X_i, that represents the detector's estimate of the likelihood that the signal was present on that trial. The list of estimates $<X_1, X_2, \ldots X_m>$ is the group estimate vector, $\mathbf{X}$. The system's task is to decide, given the group estimate vector, whether or not a signal was present.

All the noise sources are assumed to be additive, normally distributed (Gaussian) random variables having zero means and variances of σ_1^2, σ_2^2, σ_3^2, $\sigma_{1,2,3}^2$ and $\sigma_{1,3}^2$; the magnitude of the variances are independent of which stimulus event occurred. Thus, the statistic, X_i, is a normally distributed (Gaussian) random variable, having a mean of zero on noise trials and a mean of μ_i on signal trials. The difference between the means of $\mathbf{X}$, given signal and given no-signal, is the mean vector, $\mu = <\mu_1, \mu_2, \ldots \mu_m>$.

The variance of X_i is equal to the sum of the variance of its noise inputs. For detector 1 we have

$$\text{Var}(X_1) = \sigma_1^2 + \sigma_{1,2,3}^2 + \sigma_{1,3}^2 \tag{3}$$

The covariance of the estimates of any pair of detectors, $\text{Cov}(X_i, X_j)$, is equal to the sum of the variances of the noise sources common to those two detectors. For detectors 1 and 3:
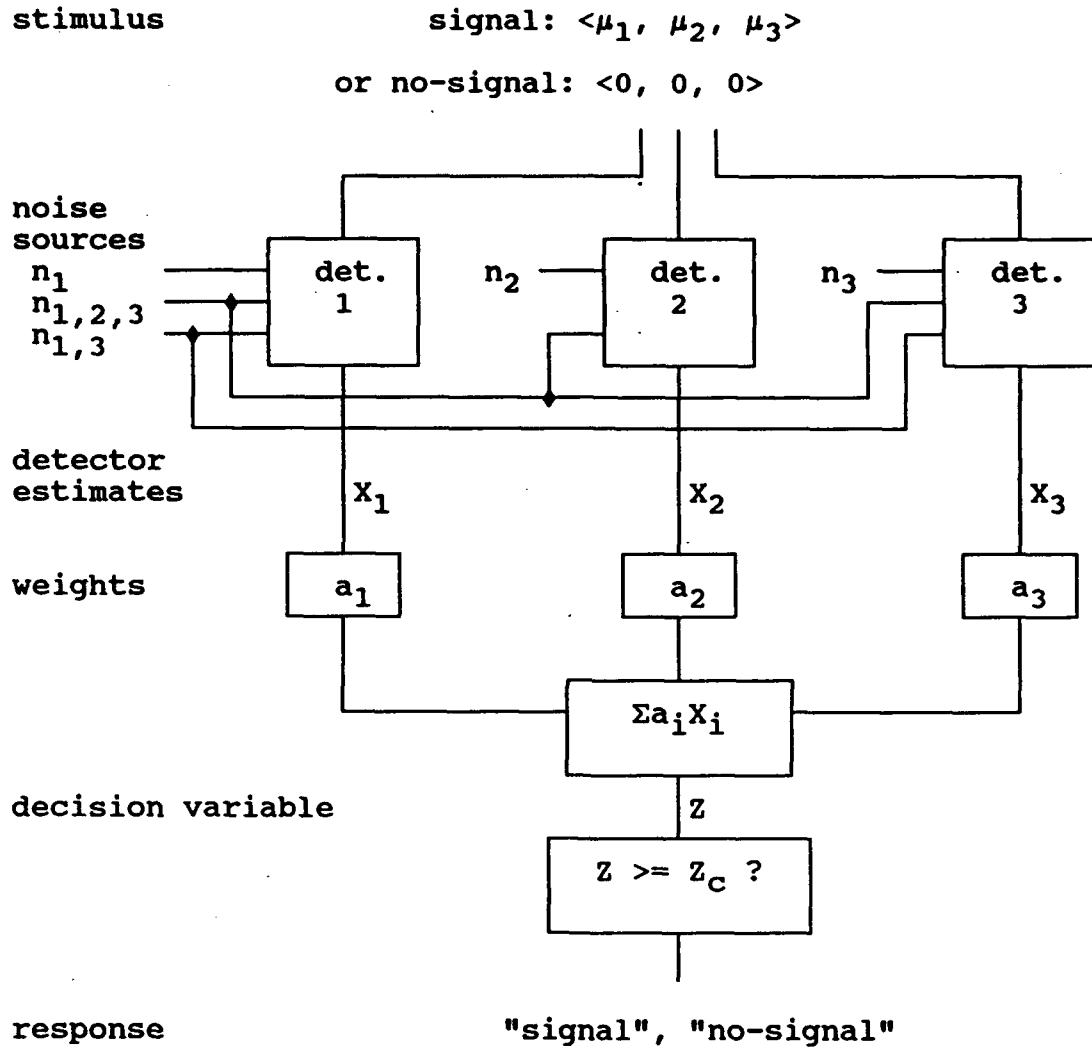
$$\text{Cov}(X_1, X_3) = \sigma_{1,2,3}^2 + \sigma_{1,3}^2 \tag{4}$$

The entries of the covariance matrix, Σ, summarize the values of these variances and covariances. For the **specific** system shown in figure 5, we have

$$\Sigma = \begin{bmatrix} \sigma_1^2 + \sigma_{1,2,3}^2 + \sigma_{1,3}^2 & \sigma_{1,2,3}^2 & \sigma_{1,2,3}^2 + \sigma_{1,3}^2 \\ \sigma_{1,2,3}^2 & \sigma_2^2 + \sigma_{1,2,3}^2 & \sigma_{1,2,3}^2 \\ \sigma_{1,2,3}^2 + \sigma_{1,3}^2 & \sigma_{1,2,3}^2 & \sigma_3^2 + \sigma_{1,2,3}^2 + \sigma_{1,3}^2 \end{bmatrix} \tag{5}$$

In the psychoacoustics literature, this detection task is framed as the problem of detecting a brandband stimulus that has components in m channels, where the channels are defined in terms of spectral, spatial, or temporal dimensions. The multi-channel auditory signal detection problem has been discussed by Berg (1989, 1990), Berg and Green (1990), Durlach et al. (1986), and

Figure 7. A simplified version of a group detection system. Each detector has a unique source of noise, plus a noise that is shared with one or both of the other detectors. The noise sources are independent, Gaussian random variables, with zero means and specified variances; the variances are independent of which stimulus event occurred. The decision variable, Z, is the weighted sum of the detector estimates.

stimulus   signal: $\langle \mu_1, \mu_2, \mu_3 \rangle$

                 or no-signal: $\langle 0, 0, 0 \rangle$

noise sources

n_1

$n_{1,2,3}$

$n_{1,3}$

| det. 1 | n_2 | det. 2 | n_3 | det. 3 |

detector estimates X_1 X_2 X_3

weights a_1 a_2 a_3

$\Sigma a_i X_i$

decision variable   Z

$Z \geq Z_c$?

response   "signal", "no-signal"

Green (1988, 1992). Note that the task also can be framed as a group detection problem, in which a group or team of detectors make the m observations and must arrive at a decision about the existence of signal.

An optimal detector employs a decision variable, Z, that is a monotonic function of the likelihood ratio statistic. As long as the covariance matrix has the same form for the signal and no-signal distributions, an optimal decision variable is a linearly weighted sum of the detector estimates (Ashby and Maddox, 1992), i.e.,

$$Z = X' \, \Sigma^{-1} \, \mu + k \tag{6}$$

where X' is a row vector, Σ^{-1} is the inverse of the covariance matrix, μ is a column vector, and k is a constant. Let the vector, $a = \Sigma^{-1} \mu$, then an equivalent decision variable is

$$Z = \sum_{i=1}^{m} a_i \, X_i \tag{7}$$

where the a_i are optimal weights applied to the estimates, X_i. The optimal weights are expressed in terms of the inverse of the covariance matrix and the mean vector. The index of detectability, $d'_{Ideal\ Group}$, for this system is (Mahalanobis, 1931):

$$d'_{Ideal\ Group} = [\, \mu' \, \Sigma^{-1} \, \mu \,]^{\frac{1}{2}} \tag{8}$$

where μ' is a row vector.

Suppose that two sources of noise enter each detector, one having a variance of σ^2_{com}, which is common to all the detectors, and the other having a variance of σ^2_i, which is unique to each detector. All of the off-diagonal elements of the covariance matrix are equal to σ^2_{com}. The optimal weights, a_i, for this case are (Durlach et al., 1986):

$$a_i = \mu_i \left(\frac{1}{\sigma^2_i} - \frac{D}{\sigma^4_i} \right) - \sum_{j \ne i}^{n} \frac{D\mu_j}{\sigma^2_j \sigma^2_i} \tag{9}$$

$$\text{where } D = \sigma^2_{com} \left(1 + \sigma^2_{com} \sum_{i=1}^{m} \frac{1}{\sigma^2_i} \right)^{-1} \tag{10}$$

The detectability index, d', (Durlach et al., 1986) is:

$$d'_{Ideal\ Group} = \left[\sum_{i=1}^{m} \left(\frac{\mu_i}{\sigma^2_i}\right)^2 - D \left(\sum_{i=1}^{m} \frac{\mu_i}{\sigma^2_i}\right)^2 \right]^{\frac{1}{2}} \tag{11}$$

The equation can be simplified further by assuming that the unique variance components are equal in magnitude across the detector array, that is

$$\sigma^2_i = \sigma^2_{ind}, \quad \text{for all } i \tag{12}$$

By definition, the correlation between any pair of detectors is given by:

$$r = \sigma^2_{com} / (\sigma^2_{ind} + \sigma^2_{com}) \tag{13}$$

Because the magnitude of the unique and common variances are uniform over the array of detectors, the detectabilities of the

individual detectors, d_i', are characterized only by the values of μ_i. We can normalize each detector's total variance by setting $\sigma_{ind}^2 + \sigma_{com}^2 = 1$, then

$$d_i' = \mu_i \ / \ (\sigma_{ind}^2 + \sigma_{com}^2)^{\frac{1}{2}} = \mu_i \qquad (14)$$

Then we have the important relationship:

$$d'_{Ideal\ Group} = \left\{ \frac{m\ Var(d')}{1 - r} + \frac{m\ (\overline{d}')^2}{1 - r + mr} \right\}^{\frac{1}{2}} \qquad (15)$$

where $\overline{d}'$ is the mean of the individual d's, Var(d') is the variance of the individual d's, m is the group size, and r is the inter-detector correlation.

As part of his masters thesis project, Chris Hays and I have begun to run experiments with human subjects to test the basic predictions of the Ideal model. Initially, we are running groups of from 2 to 8 observers in a signal detection task. After each observer is presented with a (partially correlated or uncorrelated) observation, a random member of the group is asked to give the group decision about the presence of signal or noise on that trial. Interaction is completely free, although there is a time constraint on answering. We will evaluate the efficiency of group and individual decision making in this task, as well as compute the weight of each group members input on the group decision.

1.2.2 Optimal Binary Groups (Sorkin and Dai, 1993; Sorkin and Dai, submitted)

We have begun to study the performance of arrays of detectors when the outputs of the detectors are binary in nature. Given knowledge of the group members' individual d's and criteria, a group "supervisor" could compute exactly the likelihood of signal and noise given each possible binary pattern obtainable in the group members' response array. These could be ordered in terms of likelihood ratio, and appropriate responses made to particular patterns.

In general, the particular value of group hit and false alarm rate would depend on the supervisor's criterion--as well as on the criteria of the individual detectors. We estimated the group d's obtainable under some simple assumptions about the interdetector correlation and d' and c statistics. Let $r = 0$ and $var(d') = 0$ (and $d'_i = 1$). Further suppose that all detectors employ the same individual response criterion, c. Now, all the information in the group binary response pattern is given by the number of detectors voting "yes". We can fix c and examine the hits and fas obtainable from varying the number of yeses needed for a group yes response. Likewise, we can fix the majority rule and examine the hits and fas obtainable by varying the value of c. In both cases, we obtained ROC curves which resemble normal-normal ROC curves. (Clearly, all the curves must go thru 0,0 and

24

1,1.) The highest group performance is obtained with a majority rule of 0.5.

Note that in the absence of a group strategy to weight by d', any variance in member d' will increase the difference between the majority strategy and the ideal. We believe that human groups have evolved various strategies for acquiring information about the d's and c's of the individual members. A notable example is as the jury deliberation system. We have developed some models of such groups, using iterative polling and shifting of the individual detector criteria. We think that these models are of potential interest to a range of detection systems, including those involving arrays of neurons and arrays of jurors, and alarm systems, where decisions may depend on arrays of binary outputs and where adjustments may be made in the individual detector criteria for firing.

1.2.3 Limited interaction groups (Sorkin and Crandall, Sorkin and Dai).

Groups vary in the degree of interaction among group members that occurs during deliberation. At one extreme is the hypothetical Ideal Group, in which it is assumed that members freely discuss all matters relevant to communicating the values of X_i and μ_i, and then put this information into a form appropriate for calculation of the optimum response. The other extreme is the group with no interaction among members; the members of this group simply make their private observations and then take a single vote. In between these two extremes are real groups such as committees, juries, and teams, where customs or formal rules dictate how group members communicate and how member judgments are combined to form the group decision.

One type of formally limited group interaction consists of an iterative series of ballots and discussions, such as occurs in an American jury. The group has a discussion, takes an open ballot consisting of the binary responses of each member, and counts the resulting votes. This sequence is repeated until a specified majority vote is reached, or until a time limit is exceeded.

In terms of detection theory, the group operates as follows: As a consequence of observing the stimulus evidence and prior to interaction as a group, each member makes an estimate, X_i, of the likelihood of signal. This estimate leads to a vote, R_i, of either signal, S, or no-signal, NS. The vote is based on the value of the member's observation, X_i, and the member's pre-deliberation criterion, c_i. The votes are tallied and, if unanimity is not reached, the group proceeds to discussion. During the discussion, each member acquires information about every other member's response, as well as about every each other's detectability, d_i', and criterion, c_i. Each member then uses that information to compute a new criterion. Thus, each member shifts his or her own criterion as a function of the response (R_i), the estimated detectability (d_i'), and the bias (c_i), of the other

25

team members. After a new criterion is computed, the member's original observation, X_i, is again compared with it, and a new response is made. This process is repeated until a decision or time deadline is reached. This process may be characterized as as a fixed-rule, dynamic network.

The rule for shifting a member's criterion follows from an analysis of aided detection described by Robinson and Sorkin (1985), Sorkin and Woods (1985), and Murrell (1977). An example of this system is the case of two detectors, one is a human detector and the other is an auxiliary "alarm" detector. These detectors operate together to perform a detection task. The human detector incorporates the binary response of the alarm detector to decide whether a signal or no-signal event has occurred.

According to Robinson and Sorkin (1985), the human detector incorporates the alarm detector's output by employing two different response criteria, depending on whether the alarm detector has responded signal (S) or no-signal (NS). These contingent criteria are computed using the following formula:

$$\beta \text{ (given output R from alarm detector)} = V \cdot \frac{p(ns)}{p(s)} \cdot \frac{p(R|ns)}{p(R|s)} \qquad (16)$$

where p(s) and p(ns) are the prior probabilities of signal and no-signal, respectively, and $p(R|s)$ and $p(R|ns)$ are the probabilities that the alarm detector has made response R, given signal or given no-signal, respectively. V is the ratio of payoffs to the human detector for the four possible event outcomes:

$$V = [v(NS \cdot ns) - v(S \cdot ns)] / [v(S \cdot s) - v(NS \cdot s)] \qquad (17)$$

where $v(S \cdot s)$ is the payoff for correctly-decide-signal, $v(S \cdot ns)$ is for incorrectly-decide-signal, $v(NS \cdot ns)$ is correctly-decide-no-signal, and $v(NS \cdot s)$ is incorrectly-decide-no-signal.

Equation 16 is based on the principle that the human detector should compute the posterior probability of S (and NS) given the alarm detector's response, and assumes that the human wishes to maximize expected value. That is, after receiving information about the alarm response, the human detector updates her prior probability by substituting the posterior probability based on the alarm detector's response. This updated prior probability is employed in recalculating the human detector's criterion. Note that in order to calculate $p(R|s)$ and $p(R|ns)$, it is necessary for the human detector to know the d' and criterion of the alarm detector.

If there are m independent alarm detectors,

$$\beta = V \cdot \frac{p(ns)}{p(s)} \cdot \frac{p(R_1|ns)}{p(R_1|s)} \cdot \frac{p(R_2|ns)}{p(R_2|s)} \cdot \ldots \frac{p(R_m|ns)}{p(R_m|s)} \qquad (18)$$

where R_i is the response of alarm detector i.

The team situation is much more complex than the alarm detector paradigm because, (1) each detector's output goes to all the other detectors, (2) the system decision is based on the outputs of all of the detectors rather than just the one (the human's), and (3) the system decision is dynamic--the set of detector responses changes over time as each detector modifies its decision to accommodate the influence of the others.

We have implemented this dynamic network algorithm in a computer simulation of team decision making. The most obvious group behavior produced by this algorithm is the tendency for the number of votes favoring the majority position to increase during deliberation. This occurs because a preponderance of say, S votes, shifts the average member's criterion toward making an S response. Responses from members having higher d's produce more criterion shift than responses from less sensitive members, and a member's S vote that was made using a lax criterion for S counts less than one that was made using a very strict criterion.

We can summarize some qualitative aspects of the model simulations that we have run so far. First, on most trials the algorithm results in a decision toward the position initially favored by a majority of members. Second, sometimes members' criteria oscillate over successive ballots. Third, occasionally there is a reversal of the initial majority vote. Fourth, on occasional trials a decision is not reached by the time our arbitrary stopping point is reached. These qualitative aspects of the model's behavior during group deliberation are consistent with those found in previous empirical studies and simulations, for example, by Kalven and Zeisel's (1966) study of the American jury, and of small group studies described by Saks (1977) and Penrod and Hastie (1980).

In order to perform the criterion-shift calculations required by the contingent criterion model, each team member must know the vote, detectability, and criterion of each of the other members. In some groups, limitations on member communication prevent members from acquiring this information. One group of this type is the Delphi Technique Group (Hastie, 1986; Gustafson et al., 1973), in which efforts are made to maintain the anonymity of members in order to prevent undue influence or the suppression of discussion by group members holding positions of authority. After balloting, each member is provided only with an aggregate vote that shows the number voting S and NS; no information is provided about individual d_i and c_i. It is easy to add such informational constraints to a limited interaction version of the contingent criterion model. Because specific information about the other members is not available, each member must use an _average_ estimate for the sensitivity and criterion of

27

the members giving the number of S and NS votes. Thus, calculations of p(S|s) and p(S|ns) are based on the group member's estimate of the average d' and criterion for the rest of the group. As in the Contingent Criterion case, a preponderance of S votes tends to shift members' criteria toward making an S response more likely.

Figure 8 illustrates the results of some simulations of different types of groups using different decision rules. The figure is a plot of group d' versus the size of the group. From best to worst performance, the different groups are: Ideal Group, Contingent Criterion Group-unanimous decision, Contingent Criterion Group-3/4 majority, Contingent Criterion Group-2/3 majority, Delphi group-2/3 majority, and Single Ballot-2/3 majority. All groups were assumed to have an inter-member correlation of 0, and the same distributions of member d' and β. Substantially the same results occur when the intermember correlation is greater than zero, but the differences are smaller.

We were concerned about use of the d' measure for characterizing the performance of these complex group detection systems. If the variance of the hypothesis distributions were not approximately equal, d' would not be an adequate measure, particularly for $\beta \ll 1$ or $\beta \gg 1$. Metz and Shen (1992) analyzed group detection without the requirement for the equal variance assumption. They predicted the accuracy gain in reading diagnostic images, such as X-films, that result from replicated readings by the same or different readers (all judgments were rated equally). Rather than computing a group d', they showed how the parameters of the general binormal Receiver Operating Characteristic depend on the number of readings and the within-reader and between-reader variation.

To check on the equal variance assumption for our models, we plotted the group hit and false alarm probabilities that were obtained in several conditions of simulations using different values of mean β, on Receiver Operating Characteristic (ROC) curves [P(S|s) versus P(S|ns)]. The resultant curves were quite similar to equal-variance, single-detector ROC curves. Thus, at least under the conditions evaluated by our simulations and proposed for the human experiments, the use of the d' and β measures appears to be appropriate for summarizing the performance of group systems.
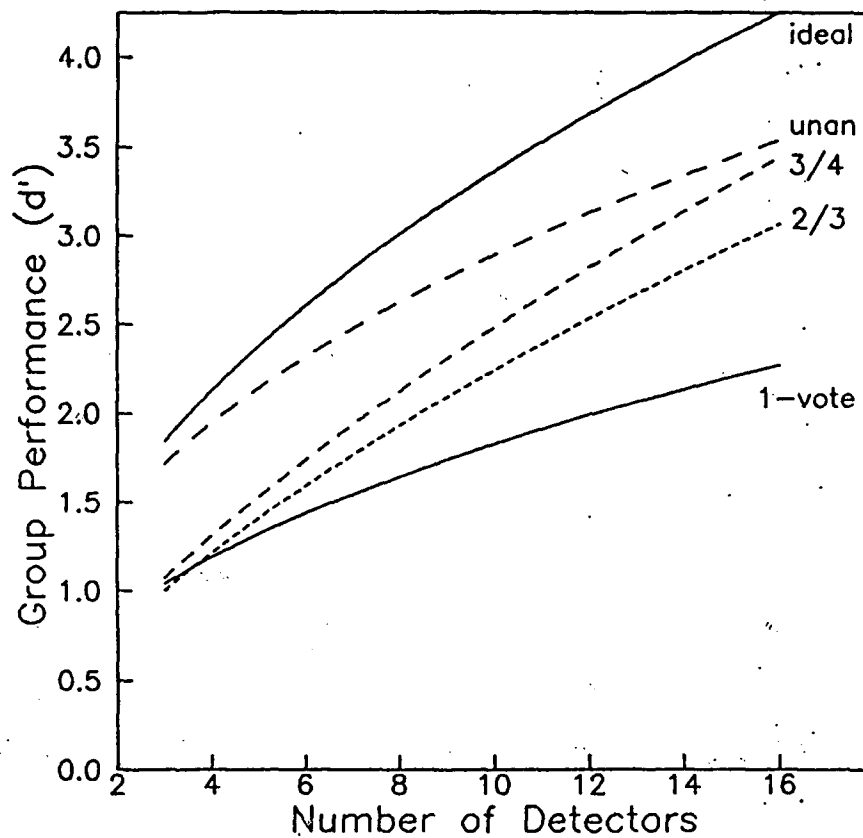
Figure 8. The performance of five different groups as a function of the group size. The group parameters are: $r=0$, $\mu_{d'}=1$, $\sigma_{d'}=0.36$, $\mu_{crit}=1$, and $\sigma_{crit}=1$.

1.3 Signal detection with multi-element displays

In these experiments, we studied an observers' ability to use multiple independent sources of visual information in a signal detection task. The objectives were to determine the

observer's efficiency at using information from different spatial locations of the display and to determine the effects of display coding and arrangement.

1.3.1 Observer sensitivity to element reliability (Montgomery and Sorkin, submitted).

Visual displays are commonly used to convey system information, such as air traffic flow or the status of a production line, to a human decision maker. A complex visual display may include several subordinate displays or display "elements." Each display element provides a potential source of information for the human operator. However, it may be impossible for the operator to obtain useful information from more than a few of the display elements at one time. This problem may be minimized if the operator can prioritize the display elements in terms of their criticality and informativeness, and if the operator can allocate his or her attention accordingly. This study examined several factors that affect an operator's ability to allocate attention to display elements that are differentially informative.

In a previous experiment (Sorkin, Mabry, Weldon, & Elvers, 1991), observers examined a multi-element display and then reported whether the display represented the occurrence of a signal or nonsignal event. Using a technique derived from the Theory of Signal Detectability (TSD, Green & Swets, 1964), Sorkin et al. estimated the importance or weight the observer assigned to each element of the display in making a detection decision (Berg, 1989, 1990).   An optimal decision-theoretic observer weights the input from each element according to the element's informativeness or reliability; highly reliable display elements are weighted more highly in the detection decision than less reliable elements (Durlach, Braida, & Ito, 1986; Berg, 1990; Berg & Green, 1990). Berg (1990) developed a measure, weighting efficiency, for assessing how accurately an observer weights display elements by their differing reliabilities.

In the Sorkin et al. (1991) study all display elements were equally informative; hence, the observers should have weighted each element equally in their detection decisions. When the observation durations were long, the resulting weights were equal across the spatial array of display elements. However, when the observation durations were brief and the display coding was complex, the highest decision weights were associated with display elements in the center of the visual field, near the observer's fixation point. The lowest performance was obtained in conditions where the weighting functions were most highly peaked. Sorkin et al. (1991) concluded that, under difficult information processing conditions, an observer's allocation of attention is restricted to the central portion of the display.

This interaction between the difficulty of the task and the availability of information from different regions of the display is not surprising. A number of variables are known to affect an

observer's ability to obtain information from the elements of a complex visual display. These include the number (Perrott et al.; 1991) and spacing (Andre & Wickens, 1988) of irrelevant, or distracter, items found in the visual field, the type of display code (Boles & Wickens, 1987; Legge, Gu, & Luebker, 1989; Sanderson, Flach, Buttigieg, & Casey, 1989; Sorkin et. al., 1991), and task complexity (Williams, 1982).

When the stimulus durations in the Sorkin et al. (1991) experiment were long (more than 500 ms), all display element weights were equal, indicating that the observers could process information from all regions of the display. Since the reliability of all the elements was also equal, an equal weight strategy was optimal for that task. An important question is whether an observer can employ optimum weights when the reliabilities of the elements are not equal across the visual array. Obviously, the ability to match decision weights to the element reliability is necessary if the observer is to prioritize the display elements according to their importance to the task.

When an informational source does not provide a consistent report of an unchanging event, the source is not very reliable. For instance, if a sensor measures a specific luminance value to be x at one time and x ± n on a subsequent reading, the sensor is showing variability in its measurement. Thus, this sensor would be less reliable than one which produces a consistent measure across time. A person forming a decision based on this information should place greater weight on the more reliable source. However, evidence suggests that people tend to overrate the importance of unreliable sources (Schum, 1975). Wickens (1984) states that when people are confronted with sources which are not equally informative, they perform the task "as if" all sources were equally reliable.

In the present study, we tested whether observers could use differences in display element variability to identify the reliabilities of different sources and whether they could use this information in forming their detection decisions. In addition, we hoped to determine whether using reliability information imposed a significant amount of additional processing "overhead" on the observer, and whether selected display factors could reduce performance decrements related to this overhead.

As in Sorkin et al. (1991), the observers in the current study performed a multi-channel visual detection task. On each trial of the experiment, observers were presented with a display consisting of nine display elements. The display elements were nine vertical line-graph gauges arranged in a horizontal array (see figure 9). The values displayed on the line-graph gauges, $<x_1, x_2, \ldots, x_9>$, were determined by independent, normally distributed, random variables. On a signal trial, the values of the nine elements were selected from a distribution with a mean of μ_s and a standard deviation of σ. On a noise trial, the values were drawn from a distribution with a mean of μ_n and a standard deviation of σ, where $\mu_n < \mu_s$. The observer's task was to decide

31

whether the data displayed had been generated from the _signal_ or _noise_ distribution.

The reliability of different display elements was controlled by manipulating the variance of the distributions from which the element values were sampled: high reliability elements were sampled from distributions with low variance and low reliability elements were sampled from distributions with high variance. A high reliability source would be analogous to an instrument which showed measurements that were consistent over time, whereas a low reliability source would be analogous to an instrument whose readings varied widely over time. The variance of a display element at a particular spatial position depended on the experimental condition, but was always the same on signal and noise trials. Table 1 illustrates the mean and standard deviations for a nine element display in which odd and even elements alternate in their level of reliability. The detection performance of a hypothetical ideal observer, based on that display element is shown on the bottom row of the table (the Appendix provides details of the theory).

TABLE 1

Means and standard deviations for the nine elements where the sources alternate in reliability; the even elements have the highest reliability.

element	1	2	3	4	5	6	7	8	9
signal									
μ_s	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
σ	1.5	0.75	1.5	0.75	1.5	0.75	1.5	0.75	1.5
noise									
μ_n	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
σ	1.5	0.75	1.5	0.75	1.5	0.75	1.5	0.75	1.5
Predicted d'	0.67	1.33	0.67	1.33	0.67	1.33	0.67	1.33	0.67

We were interested in whether observers would be sensitive to the reliability of individual elements in the absence of additional cues to element reliability. That is, can observers estimate (and employ) information about element reliability based only on the variability of the readings from individual display elements and on feedback about the signal and noise events? To answer that question, we introduced conditions in which the relationship between element spatial position and reliability was either random or fixed over a block of 200 trials. In the random block condition, the spatial position of the high reliability display elements varied randomly over trials. Thus, in this condition observers could not use the trial-to-trial variability of a spatial element to identify which sources were most reliable. In the fixed block condition, however, the observer could estimate the variance of the element readings from the first k trials of a block. Using that estimate, the observer might be able to partition the elements into those with high and low reliabilities. If that process led to the assignment of higher weights to the more reliable elements, the observer's weighting efficiency would be enhanced in that condition relative to the random block condition.

A third experimental manipulation was included to test the efficacy of providing a cue to element reliability. In an auditory task similar to the one used in this study, Berg (1990) found that observers were better at weighting sources according to reliability when the most reliable tones were much louder than the less reliable tones. The loudness cue was much less effective when reversed, i.e. when a louder cue indicated a lower reliability. Berg's results suggest that under some conditions cuing element reliability (e.g., with intensity or color) may aid observers in accurately weighting display sources by their importance.

Cues such as size, intensity, color, and movement are often incorporated in display design to draw attention to specific items in a display. For instance, researchers have found that correct utilization of color coding (Christ, 1990; Fisher & Tan, 1989) can reduce search time in locating an item in a display. Furthermore, Wickens and Andre (1990) showed that color coding items in an object display lead to improved accuracy in recalling the specific value associated with a given item. Given these results, we predicted that the efficiency of the observer's weighting strategy should be higher for a condition in which a luminance cue signalled the element reliability.

In the luminance cue condition, the luminance of the display element was either high or low in accordance with the reliability (high or low) of the element. The spatial position of differentially reliable sources varied randomly across trials. We expected that luminance would provide a natural code for allocating observer attention and hence weight, to the high reliability elements. If that were the case, the efficiency of the observers' weighting strategy would be much higher in a cued than in an uncued condition.

33

Finally, the stimulus duration and the particular spatial arrangement of element reliabilities were also expected to influence observers' ability to match weights to the element reliabilities. The results from the Sorkin et al. (1991) study suggested that 233-ms was sufficient time for observers to utilize information from as many as nine, equally reliable, graphically coded display elements. However, it is possible that sensing the element reliabilities and differentially weighting the elements, could require some additional processing steps or "overhead" by an observer. A duration of 233-ms may be at the margin of an observer's ability to extract the information needed to discriminate and employ differences in element reliability. For example, a slower, serial search may be required to both extract the reliability information and to weight the information from the elements. In that case, it might be advantageous for an observer to ignore reliability, when processing short duration stimuli, and to weight all elements equally. Our experiments tested three levels of stimulus duration (150, 400 and 800-ms). We expected that weighting efficiency would be greatest at long stimulus durations (400-ms and 800-ms) and poorest at the shortest duration (150-ms).

Observer sensitivity to element reliability also may be affected by the spatial arrangement of element reliability. If attention is distributed more effectively among spatially contiguous than separated items, grouping sources similar in reliability should aid performance. Posner, Snyder, and Davidson (1980) found that simple reaction time to detect a light at a second most likely position was facilitated only when this item was adjacent to a cued location (the most likely target location). When the second most likely position was separated by more than one location, detection speed was not facilitated. Thus, weighting efficiency should be better for displays with elements grouped by similar reliabilities than for displays that distribute element reliabilities across the array.

Four University of Florida students with normal, or corrected to normal, visual acuity participated as observers in this study. One subject, S2, was later discovered to be color deficient. Another subject, S4, had extensive experience with the task. Subjects were paid an hourly wage plus a bonus based on performance.

Observers were seated in a sound isolated booth approximately 27 inches away from a 10.5 inch color monitor (EGA) driven by an 80386 computer. The monitor was set for maximum contrast, and intensity was set at approximately 102 cd/m^2, measured from a 7.5 inch by 4 inch uniform white field. On a given trial, nine gauges were presented on the monitor; subtending a horizontal by vertical visual angle of approximately 16° by 8°. Each gauge was composed of two parallel white lines, with tick-marks falling at equal intervals on the left line for all conditions except the luminance cue condition. For this latter condition high reliability gauges were white and the remaining gauges were gray. The intensity of the white gauges was approximately 102 cd/m^2 and

the intensity of the gray gauges was approximately 22 cd/m² measured from 7.5 inch by 4 inch uniform white and gray fields, respectively.

Each tick-mark represented a display increment of 1.0, and ranged from 0.0 to 10.0. Two longer blue lines, located near the tick-marks, indicated the positions of the signal and noise distribution means. The value displayed by each gauge was determined by sampling a number from either a "signal" or "noise" distribution, depending on the type of trial. This number was converted to the vertical displacement of a horizontal white line from the bottom (e.g. zero position) of the gauge (see figure 9). The gauge values were drawn from the signal distribution on 50 percent of the trials. The mean of the gauge values on signal trials, μ_s, was equal to 5.0; the mean on noise trials, μ_n, was equal to 4.0. The standard deviation of the gauge values on signal and noise trials depended on the particular condition.
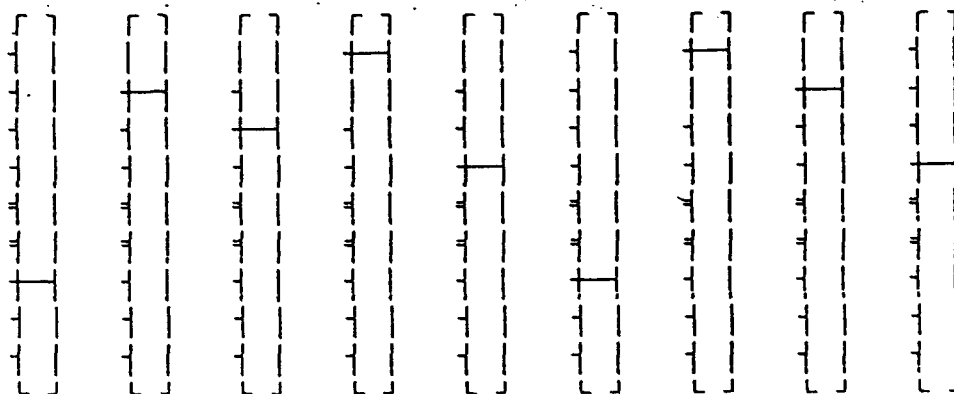
Figure 9. Example of the 9-element graphical display.

TABLE 2

Summary of experimental conditions.

X 4 subjects 3 durations (150, 400, 800 ms)			UNCUED	CUED
UNEQUAL RELIABILITY CONDITIONS	MIXED BLOCK	Grouped Left	hhhh11111	HHHH11111
		Grouped Right	11111hhhh	11111HHHH
		Distributed Even	1h1h1h1h1	1H1H1H1H1
		Distributed Odd	h1h1h1h1h	H1H1H1H1H
	PURE BLOCK	Grouped Left	hhhh11111	
		Grouped Right	11111hhhh	
		Distributed Even	1h1h1h1h1	
		Distributed Odd	h1h1h1h1h	
EQUAL RELIABILITY CONDITIONS			mmmmmmmmmm	

36

The experimental conditions are summarized in table 2. There were five different element reliability arrangements: (1) Equal, (2) Grouped-Left-High, (3) Grouped-Right-High, (4) Distributed-Even-High, and (5) Distributed-Odd-High. In the Equal condition, the standard deviation of all gauge elements was equal to 1. In the Grouped-Left-High condition, the standard deviation of the four left elements was equal to 0.75, and the five right elements was equal to 1.5. That pattern was reversed in the Grouped-Right-High condition. In the Distributed-Even-High condition, the standard deviation of the four even elements (element 2, 4, 6, and 8) was equal to 0.75, and the standard deviation of the remaining elements was equal to 1.5. In the Distributed-Odd-High condition, the standard deviation of the five odd elements (element 1, 3, 5, 7, and 9) was equal to 0.85, and the standard deviation of the remaining elements was equal to 1.3. The standard deviations were selected to maintain predicted optimal performance, d'_{ideal}, at 3.0.

The unequal reliability conditions were run under two different trial block conditions: Pure Block and Mixed Block. In the Pure Block condition, all display and distribution parameters were fixed within a block of 200 trials. Thus, for the four arrangements (Grouped-Left-High, Grouped-Right-High, Distributed-Even-High, and Distributed-Odd-High), the relationship between element reliability and spatial position was fixed throughout the block of trials. In the Mixed Block conditions, the trials within a block of 200 trials alternated randomly among the Grouped-Left-High, Grouped-Right-High, Distributed-Even-High, and Distributed-Odd-High arrangements. Therefore, in the Mixed Block conditions the reliability of an element at a given spatial position was random over trials.

All trial block conditions were tested at three levels of stimulus duration (150, 400, and 800 ms). The duration of the stimulus presentation was synchronized with the refresh traces of the monitor. The period between traces was approximately 17 ms. The onset and offset of the display was delayed until a retrace was ready to occur. Once the stimulus was presented, the duration was controlled by counting the number of refresh traces which corresponded with the selected stimulus duration (150, 400, or 800 ms).

Procedure

Observers were told to make their decisions based on the level of the gauges relative to the **signal** and **noise** mean markers. They were told to rank the likelihood that the evidence represented a signal by using the "4", "3", "2" and "1" keys, where "4" represented very sure it was a signal and "1" represented noise. In fact, observers tended only to use the two middle keys. Thus, responses on keys "1" and "2" were combined to represent noise, and responses on keys "3" and "4" responses were combined to represent signal in the data analyses. When the reliabilities differed across elements, observers were informed that the least variable gauges were the most reliable.

37

The trial sequence proceeded as follows. First, observers were given a 0.5" by 0.5" fixation cross at the center of the display for 200 ms. This was replaced by the nine line-graph gauges for a stimulus duration of either 150, 400, or 800 ms. Following the stimulus a white blanking mask was presented for 200 ms. Then, the display was completely black for 1 second, at which time the observers were allowed to respond. Any responses made prior to or following this period were discarded as "No Response" trials. Finally, the observers were given feedback at the center of the display for 250 ms. Within a given session, an observer ran through 10 blocks of 200 trials. Across sessions there were 1500 trials (750 signal and 750 noise) collected for each condition.

Due to time constraints imposed by the need to collect multiple trials, some of the observers received less practice than others. Subject S4 was highly practiced. He ran through at least eight practice sessions for each condition prior to collection of the experimental trials. Subjects S1, S2, and S3 were highly practiced on the Yes/No detection task, but they only ran through one practice session for each of the individual conditions.

Finally, to control for any possible practice effects in the experimental sessions, each observer received a different order of four trial block conditions, organized such that across subjects each condition occurred once in each the four possible positions in the order.

TABLE 3

Average performance (d') of the four observers for each condition.

			UNCUED			CUED		
		Arrangement	150	400	800	150	400	800
UNEQUAL RELIABILITY CONDITIONS	MIXED BLOCK	Grouped	1.845	2.003	2.099	1.986	2.288	2.371
		Distributed	1.996	2.119	2.229	1.966	2.131	2.181
	PURE BLOCK	Grouped	1.728	1.983	2.041			
		Distributed	1.841	2.021	2.086			
EQUAL RELIABILITY CONDITIONS			2.240	2.440	2.570			

Results

Table 3 summarizes the average performance (d'$_{obs}$ values) of the four observers for the experimental conditions. The values of d'$_{obs}$ for the Distributed-Even and -Odd arrangements and the Grouped-Left and -Right conditions were averaged, respectively, in order to create the distributed and grouped condition averages shown in the table. Thus, for the unequal reliability conditions shown in table 3, each entry represents the average of eight d'$_{obs}$ values (the values shown in the equal reliability conditions are the average of four d'$_{obs}$ values).

A repeated-measures analysis of variance was performed on the d'$_{obs}$ results for the unequal reliability conditions. Performance improved as stimulus duration increased ($F(2,6)$ = 13.663, $p<0.01$). In addition, there was a marginal advantage for the luminance cue condition over the two non-luminance cue conditions ($F(2,6)=3.846$, $p<0.1$). To compare performance in the equal and unequal reliability conditions, the data in the unequal reliability conditions were further collapsed across the two source reliability arrangements and a second analysis of variance performed. Again, there was a significant effect of stimulus duration ($F(2,6)=12.503$, $p\leq0.01$) and of block type ($F(3,9)=5.281$, $p\leq0.05$). All observers showed better performance in the equal reliability condition than in the unequal reliability conditions.

To summarize, these results indicated that stimulus duration, block type, reliability distribution, and cueing, had an effect on the accuracy of observer detection performance. It is logical to suspect that these differences in performance are related to the observers' weighting strategies. Next, we consider the effect of the experimental conditions on the observers' weighting strategies.

The observers' weights were estimated using Berg's (1989, 1990) Conditional-On-A-Single-Stimulus (COSS) analysis technique, described in detail in the Appendix. The estimated weights were based on the slopes of cumulative normal functions that had the best Chi-Square fit with the observers' COSS functions. Two weight estimates were calculated for each element in each condition, one for signal and one for noise trials. Out of the many COSS functions that we fitted (minimum Chi-square) to cumulative normal distributions in this analysis, only 6.7% differed significantly ($p<=0.05$) from normality. The signal and noise weight estimates were averaged and these average weights, a$_i$, were then used to compute measures of weighting efficiency, η_{wgt}. The weighting efficiency, η_{wgt}, provides a measure of how well the observer's weights match the weighting pattern that would be optimal for the particular element reliabilities in the task.

Table 4 summarizes the average weighting efficiencies for the observers. As with the d'$_{obs}$ analysis, the efficiency measures for the Grouped-Left and -Right and Distributed-Even and -Odd arrangements were averaged together to obtain the grouped and distributed efficiency values, respectively. In general, the

weighting efficiency results correspond to the d'_{obs} results. Across all conditions, weighting efficiency increased as the stimulus duration increased.

===

TABLE 4

Average weighting efficiency estimates for the experimental variables (arrangement, stimulus duration, block-cueing condition)

			UNCUED			CUED		
		Arrangement	150	400	800	150	400	800
UNEQUAL RELIABILITY CONDITIONS	MIXED BLOCK	Grouped	0.580	0.626	0.662	0.790	0.868	0.885
		Distributed	0.658	0.694	0.739	0.689	0.778	0.819
	PURE BLOCK	Grouped	0.589	0.703	0.707			
		Distributed	0.615	0.694	0.717			
EQUAL RELIABILITY CONDITIONS			0.750	0.848	0.868			

To identify the magnitude of statistically significant differences in η_{wgt}, a Monte Carlo simulation was performed. The parameters for the sampling distribution of η_{wgt} were chosen to provide a match to the weighting efficiency of the poorest observer. This provided the most conservative (i.e., largest) estimate of the standard deviation of the values of η_{wgt} observed in the actual experiment (estimated $\sigma_{sim} = 0.04$). The criterion that we selected for a significant difference in η_{wgt}, was two standard deviations ($2\sigma_{sim}$) of the sampling distribution of η_{wgt}. Thus, differences in η_{wgt} which exceeded 0.08 were identified as significant. Given this criterion, many of the differences in the performance accuracy shown in table 3 for different conditions, can be attributed to differences in the efficiency of the observer's weighting strategy.

Many of the differences in weighting efficiency shown in table 4 for the different conditions (block type, unequal and equal reliability, cued and uncued), can be seen to be consistent with the differences observed in the performance measures shown in table 3. For example, weighting efficiency was greatest in the equal reliability and in the cued, unequal reliability conditions. Both conditions showed an advantage over the other uncued, unequal reliability conditions. This pattern was maintained at all stimulus durations and was fairly consistent across the four observers.

Figure 10 shows the averaged weights for the two levels of source reliability and the three unequal reliability block type conditions, at a stimulus duration of 800 ms. The data are shown for the 800 ms condition because performance was highest at this duration and differences among the conditions were fairly consistent across levels of stimulus duration. Each graph represents the data from an individual observer. The weight estimates, a_i, for the separate sources were assigned to one of the two levels of reliability, depending on the variability of the stimulus values associated with that source. The data were averaged across the different types of element arrangements (left-right and even-odd); we also partitioned the data by the separate arrangements and did not find any change in the trends shown in the figure). From the figure, one can see that all the observers assigned higher weights to the high reliability than to the low reliability sources. Consistent with the analysis of weighting efficiency summarized in table 4, all observers showed the largest difference between the high and low reliability weights in the unequal, mixed block, cued condition.

Conclusions

The primary goal of this study was to determine whether observers can appropriately direct their attention to differentially informative elements of a visual display. The evidence from research on human decision making suggests that when informational sources differ in informativeness, decision makers generally do not consider these differences in forming their decisions. Instead, they act as though the sources are equally informative, and weight them accordingly (Schum, 1974; Wickens, 1984). Similarly, Berg (1990) found that observers in an auditory discrimination task were better at weighting sources that were equal rather than unequal in reliability.

In the uncued conditions of the present study, the weights that the observers' assigned to the high reliability elements were only slightly higher than the weights they assigned to the low reliability elements. Thus, observer weighting efficiencies in the uncued conditions (between 0.5 and 0.74), were generally lower than they were in the conditions when the element reliabilities were uniform. These results are consistent with earlier studies that concluded that observers tend toward using uniform weights when processing displayed information. But there are at least two exceptions to this rule. The first exception occurs

41
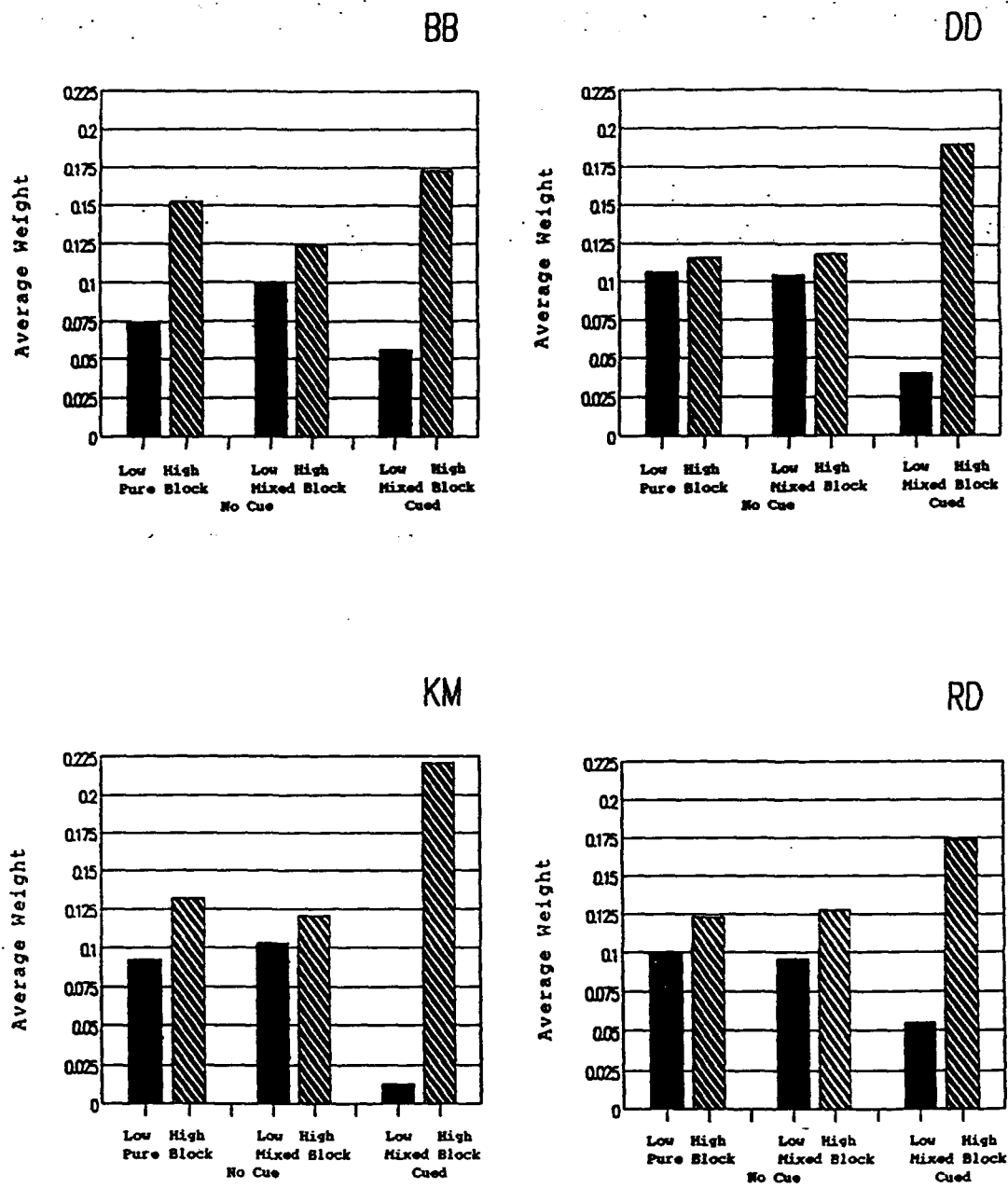
Figure 10. Average weights for four observers in the mixed-block-cued, mixed-block-uncued, and pure-block-uncued conditions as a function of the level of reliability (Low/High) at a duration of 800-ms.

when the sensory component of the task is difficult, such as when the stimulus duration is very brief. Under these conditions, the observer only can attend to a narrow display area around the fixation point; the result is a high weight for elements in the central region and a low weight elsewhere (Sorkin et al., 1991). The second exception to the rule can be seen in the present study. As the duration of the stimulus was increased, observers were able to make greater use of the differential reliability of the display elements. This was most evident when the element reliability was conveyed via a luminance cue. The weighting efficiency increased by as much as 50% when the stimulus duration was long and there was a luminance cue.

When sources have to be prioritized in terms of the underlying statistical properties of the information, observers may be limited by their ability to estimate stimulus properties such as the variability of the display elements. They may also be limited by their ability to then weight the sources appropriately, according to the estimated variability. The relatively high weighting efficiencies observed in the cued conditions of the present experiment, indicate greater observer attention to the higher reliability elements. Of course, one cannot conclude from this result that observers have improved sensitivity to the differences in element <u>variability</u>.

However, the results observed in the uncued conditions strongly suggest that observers <u>are</u> able to estimate element reliability from the statistics of the displayed information alone. Even though the weighting efficiencies in the uncued conditions were lower than those in the equal reliability condition, the observers were able to assign higher weights to the more reliable display elements. The initially surprising result was a lack of a performance advantage for the pure over the mixed block conditions. In the pure block condition, the variability of each display element was assigned to a particular spatial position and didn't change over trials. In the mixed block condition, the spatial positions of the high reliability elements changed randomly over trials. We had thought that the observers would not be able to identify the high reliability elements in the mixed block condition, because it would be impossible to estimate the variability of a given spatial position, over trials. But the results in the pure block condition suggest that the observers do not estimate the variability of specific elements over trials.

If the observers don't use information about the trial-to-trial variability of different display elements, how do they obtain information about the differential reliability of elements? It appears that observers are able to utilize variability information that is present <u>within</u> a single trial. Consider that on a given trial the readings of the high reliability elements will tend to fall at a common vertical position in the display, causing them to line up as shown in figure 4. Thus, a tighter pattern of the data displayed by the high reliability elements, provides potential information which the observer may use to

43

identify which sources were more reliable. After the experiment, we questioned the observers about strategies they used on these conditions, and some reported that they had used such display patterns in their decision making. Apparently, observers can utilize information about the relative variability of different display elements, from the within-trial pattern of displayed information.

From this study, we may conclude that observers are able to obtain information about the reliability of different display elements, but that they are are relatively inefficient at doing so. One means by which observers may estimate the reliability of different display elements is via the variability of subordinate, display patterns. However, observers show greatly improved efficiency when the display elements are coded by luminance. Appropriately designed luminance cues, and possibly other cues, can greatly help observer's to prioritize the information in a display, by indicating where attention should be directed.

Grouped-Left High
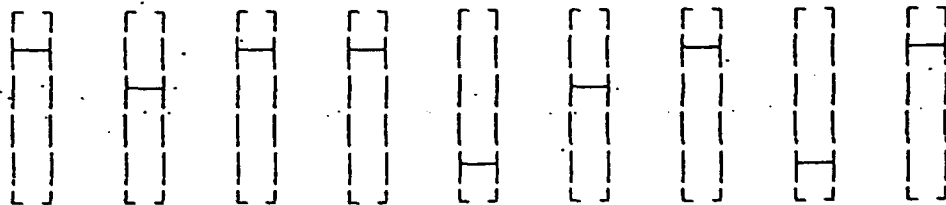
Distributed-Even High
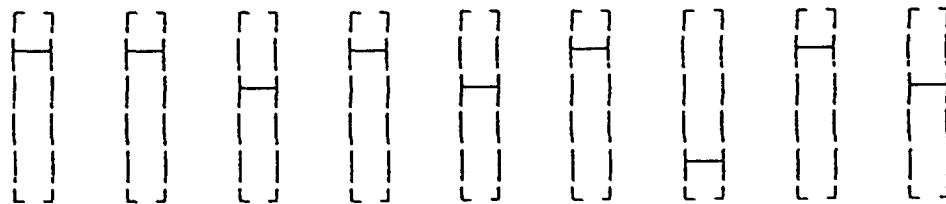
Figure 11. Example of a pattern that results from data displayed by High (and Low) reliability elements on a given trial, for the grouped and distributed arrangement.

1.3.2 Optimized Codes for visual display processing (Montgomery and Sorkin).

These experiments studied observers' ability to use multiple independent visual information sources in forming a decision. The goal of the study was to identify means of coding the (independent) visual elements so as to maximize the efficiency of decision making. The information provided by a given source is a quantity that changes in magnitude depending on the underlying state, signal or noise. As with the previous study, this quantity was represented as the value of a graphical element in a visual display. We examined the effects of two specific factors on an observer's ability to use the information conveyed by the separate elements. The first factor was whether or not the arrangement of elements produces an emergent, object-like feature. The second factor is the relationship between the emergent feature and the optimal decision statistic for the task. These experiments were reported in Montgomery and Sorkin, 1993 and in Montgomery, 1993 (attached).

2 REFERENCES

Ashby, F. G. and Maddox, W. T. (1992). Complex decision rules in categorization: contrasting novice and experienced performance. Journal of Experimental Psychology: Human Perception and Performance, 18, 50-71.

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. Journal of the Acoustical Society of America, 86, 1743-1746.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. Journal of the Acoustical Society of America, 88, 149-158.

Berg, B. G. and Green, D. M. (1990). Spectral weights in profile listening. Journal of the Acoustical Society of America, 88, 758-766.

Boles, D. B. & Wickens, C. D. (1987). Display formatting in information integration and nonintegration tasks. Human Factors, 29, 395-406.

Burgess, A. and Barlow, H. B. (1993). The precision of numerosity discrimination in arrays of random dots. Vision Research, 23, 811-820.

Buus, S. (1985). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am., 78, 1958-1965.

Cohen, M. F., and Schubert, E. D. (1987). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am., 81, 721-723.

Collins, L. M. and Wakefield, G. H. (1992) "Temporal pattern discrimination and speech recognition under electrical stimulation," Association for Research in Otolaryngology, Tampa, FL.

Davis, J. H., Kerr, N. L., Atkin, R. S., Holt, R., & Meek, D. The decision process of 6- and 12-person mock juries assigned unanimous and two-thirds majority rules. Journal of

Personality and Social Psychology, 1975, 32, 1-14.

Durlach, N. I., Braida, L. D., and Ito, Y. (1986). Toward a model for discrimination of broadband signals. Journal of the Acoustical Society of America, 80, 63-72.

Eriksen, C. W. & Rohrbaugh J. (1970). Visual masking in multielement displays. Journal of Experimental Psychology: General, 83, 147-154.

Espinoza-Varas, B. and Watson, C. S. (1986). "Temporal discrimination for single components of nonspeech auditory patterns," J. Acoust. Soc. Am., 80, 1685-1694.

Green, (1992). The number of components in profile analysis tasks. Journal of the Acoustical Society of America, 91, 1616-1623.

Gustafson, D. H., Shukla, R. K., Delbecq, A., & Walster, G. W. (1973). A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. Organizational Behavior and Human Performance, 9, 280-291.

Hall, J. W., III, Cokely, J. A., and Grose, J. H. (1988). "Combined monaural and binaual masking release," J. Acoust. Soc. Am., 83, 1839-1845.

Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am., 76, 50-56.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), Decision research (Vol. 2, pp 129-157). Greenwich, CT: JAI Press.

Hirsh, I. J., Monahan, C. B., Grant, K. W., and Singh, P. G. (1990). "Studies in auditory timing: 1. Simple patterns," Perception & Psychophysics, 47, 215-226.

Kalven, H., Jr., & Zeisel, H. The American Jury. Boston: Little, Brown, 1966.

Kidd, G. R. and Watson, C. S. (in press). "The 'proportion-of-total-duration rule' for the discrimination of auditory patterns," J. Acoust. Soc. Am.

Legge, G. E., Gu, Y. & Luebker, A. (1989). Efficiency of graphical perception. Perception and Psychophysics, 46, 365-374.

Mahalanobis, P. C. (1931). On the generalized distance in statistics. Proc. Natl. Inst. Sci. Calcutta, 12, 49-55.

Metz, C. E. and Shen, J. (1992). Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of ROC analysis. Medical Decision Making, 12, 60-75.

Montgomery, Demaris A. The effects of display coding factors on observer visual signal detection. Ph.D. Dissertation, University of Florida, 1993.

Montgomery, D. A. and Sorkin, R. D. The effects of display code and its relation to the optimal decision statistic in visual signal detection. Proceedings of the Human Factors and Ergonomics Society, Vol. 2, 1993, 1325-1329.

Montgomery, D. A. and Sorkin, R. D. Observer sensitivity to the source reliability in a multi-element visual display. Human Factors, (submitted).

Moore, B. C. J., and Emmerich, D. S. (1990). "Monaural envelope correpation perception, revisited: Effects of bandwidth, frequency separation, duration, and relative level of the noise bands," J. Acoust. Soc. Am., 87, 2628-2633.

Moore, B. C. J., and Schooneveldt, G. P. (1990). "Comodulation masking release as a function of bandwidth and time delay between on-frequency and flanking-band maskers,"J. Acoust. Soc. Am., 88, 725-731.

Murrell, G. W. (1977). Combination of evidence in perceptual judgment. In M. F. Kaplan & S. Schwartz (Eds.) Human judgment and decision processes in applied settings, (pp. 169-201). New York: Academic.

Penrod, S. & Hastie, R. (1980). A computer simulation of jury decision making. Psychological Review, 87, 133-159.

Pomerantz, J. R., (1981). Perceptual organization in information processing. In M. Kubovy and J. R. Pomerantz (Eds.), Perceptual organization (pp.141-180). Hillsdale, NJ: Erlbaum.

Povell, D. J., and Essens, P. (1985). Perception of temporal patterns. Music Perception, 2, 411-440.

Richards, V. M. (1987). "Monaural envelope correlation perception," J. Acoust. Soc. Am., 82, 1621-1630.

Richards, V. M. (1988). "Components of monaural envelope correlation perception," Hear. Res., 35, 47-58.

Robinson, D. E. & Sorkin, R. D. (1985). A Contingent Criterion Model of Computer Assisted Detection. In R. Eberts & C. G. Eberts (Eds.) Trends in Ergonomics/Human Factors, Vol. II. North-Holland: Amsterdam, 75-82.

Sadralodabai, T. and Sorkin, R. D. Discrimination of tonal sequences composed of repeated temporal patterns. Journal of the Acoustical Society of America, 1992, 91, 2333 (abstract).

Sadralodabai, T., and Sorkin, R. D. Serial position effects in temporal pattern discrimination. Journal of the Acoustical Society of America, 1993, 93, 2385-2386 (abstract).

Saks, M. (1977). Jury verdicts: The role of group size and social decision rule. Lexington, MA: Lexington.

Sanderson, P. M., Flach, J. M., Buttigieg, M. S., & Casey, E. J. (1989). Object displays do not always support better integration task performance. Human Factors, 31, 183-187.

Sorkin, R. D. (1987). "Temporal factors in the discrimination of tonal sequences," J. Acoust. Soc. Am., 82, 1218-1226.

Sorkin, R. D. (1990). "Perception of temporal patterns defined by tonal sequences," J. Acoust. Soc. Am., 87, 1695-1701.

Sorkin, R. D. and Crandall, C. S. A Signal Detection Model of the Jury. Psychological Review, (distributed for comments).

Sorkin, R. D. and Dai, H. Signal Detection Analysis of the Ideal Group. <u>Organizational Behavior and Human Decision Processes</u> (in press).

Sorkin, R. D. and Dai, H. Psychoacoustic models of group signal detection. <u>Journal of the Acoustical Society of America</u>, 1993, <u>93</u>, 2366 (abstract).

Sorkin, R. D., Mabry, T. R., Weldon, M., and Elvers, G. (1991). Integration of information from multiple element displays. <u>Organizational Behavior and Human Decision Processes</u>, <u>49</u>, 167-187.

Sorkin, R. D. and Montgomery, D. A. (1991). "Effect of time compression and expansion on the discrimination of tonal patterns," <u>J. Acoust. Soc. Am., 90</u>, 846-857.

Sorkin, R. D., Montgomery, D. A., and Sadralodabai, T. Effect of inter-sequence interval on the discrimination of tonal patterns. <u>Journal of the Acoustical Society of America</u>, (accepted for publication).

Sorkin, R. D., Robinson, Γ and Berg, B. G. (1987). A detection theory method for the analysis of visual and auditory displays. <u>Proceedings of the Human Factors Society</u> (31st Annual Meeting (Vol.2), (pp.1184-1187).

Sorkin, R. D. & Woods, D. D. Systems with Human Monitors: A Signal Detection Analysis. <u>Human-Computer Interaction</u>, 1985, <u>1</u>, 49-75.

Treisman, A. (1986). Properties, parts, and objects. In K. Boff, L. Kaufman, and J. Thomas (Eds.), <u>Handbook of perception and human performance: Vol.II. Cognitive processes and performance</u> (pp.35-70). New York: Wiley.

Watson, C. S., Foyle, D. C., and Kidd, G. R. (1990). Limits of pattern discrimination for patterns with various durations and numbers of components. <u>J. Acoust. Soc. Am.</u>, <u>88</u>, 2631-2638.

Wickens, C. D. and Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. <u>Human Factors</u>, <u>32</u>, 61-77.

3 PUBLICATIONS

Yost, W. A., Braida, L. D., Hartmann, W. M., Kidd, G. D. Jr., Kruskal, J. B., Pastore, R. E., Sachs, M. B., Sorkin, R. D., & Warren, R. M. <u>Classification of Complex Nonspeech Sounds</u>, Committee on Hearing, Bioacoustics, and Biomechanics, National Research Council, National Academy Press, Washington, DC, 1989; and <u>Journal of the Acoustical Society of America</u>, 1992, <u>91</u>, 537-8(B).

Sorkin, R. D. Perception of temporal patterns defined by tonal sequences. <u>Journal of the Acoustical Society of America</u>, 1990, <u>87</u>, 1695-1701.

Sorkin, R. D., Mabry, T. R., Weldon, M., and Elvers, G. Integration of information from multiple element displays. <u>Organizational Behavior and Human Decision Processes</u>, 1991, <u>49</u>, 167-187.

Sorkin, R. D. and Montgomery, D. A. Effect of time compression and expansion on the discrimination of tonal patterns. <u>Journal of the Acoustical Society of America</u>, 1991, <u>90</u>, 846-857.

Montgomery, D. A. and Sorkin, R. D. The effects of display code and its relation to the optimal decision statistic in visual signal detection. <u>Proceedings of the Human Factors and Ergonomics Society</u>, Vol. 2, 1993, 1325-1329.

Montgomery, Demaris A. The effects of display coding factors on observer visual signal detection. Ph.D. Dissertation, University of Florida, 1993.

Sorkin, R. D. and Dai, H. Signal Detection Analysis of the Ideal Group. <u>Organizational Behavior and Human Decision Processes</u> (in press, scheduled for Feb. 1995).

Montgomery, D. A. and Sorkin, R. D. Observer sensitivity to the source reliability in a multi-element visual display. <u>Human Factors</u>, (submitted).

Sorkin, R. D., Montgomery, D. A., and Sadralodabai, T. Effect of inter-sequence interval on the discrimination of tonal patterns. <u>Journal of the Acoustical Society of America</u>, (accepted for publication).

<u>In preparation</u>:

Sorkin, R. D. and Dai, H. Detection theory analyis of a human-machine detection system. <u>Proceedings of the Human Factors and Ergonomics Society</u>, (submitted).

Sorkin, R. D. and Crandall, C. S. A Signal Detection Model of the Jury. <u>Psychological Review</u>, (distributed for comments).

Sadralodabai, T. and Sorkin, R. D. Effect of repetition on the discrimination of tonal patterns (in preparation).

Sorkin, R. D. and Hays, C. J. *Detection theory analysis of small groups: Deviations from the Ideal Group.* <u>Journal of Experimental Psychology: Applied</u>, (in preparation).

Kantowitz, B. K. and Sorkin, R. D. <u>Human Factors</u>, 2nd edition, J. Wiley & Sons, (in preparation).

4 PERSONNEL ASSOCIATED WITH THE RESEARCH PROJECT

Crandall, Christian S., Assistant Professor of Psychology, University of Kansas. Professor Crandall assisted with the group detection study.

Dai, Huanping, Assistant Scientist, Hearing Research Center, Department of Psychology. Dr. Dai has been assisting with the analysis of Ideal Groups.

Hays, Christopher J. Graduate Student, Department of Psychology, University of Florida. Mr. Hays is a first year graduate student on assignment from the Air Force to study for a MS in Psychology. He is assisting with the group detection project.

Montgomery, D. A. Graduate Student, Department of Psychology, University of Florida. Ms. Montgomery has been on the project since May, 1989. She completed her Ph.D. in the Spring of 1993 and is an assistant professor of psychology at Illinois State University.

Sadralodabai, Toktam. Graduate Student, Department of Psychology, University of Florida. Ms. Sadralodabai joined the project in August, 1991, and has been working on the rhymicity experiments.

Sorkin, R. D. Principal Investigator, Professor of Psychology, University of Florida.

Walzak, Andrea. Graduate Student, Department of Psychology, University of Florida. Ms. Walzak assisted in the project's laboratory during the Fall 1993 semester.

5 INTERACTIONS

5.1 Papers presented at meetings, conferences, and seminars.

Sorkin, R. D. "Signal Detection Models of Group Decision Making", Department of Psychology, University of Central Florida, Orlando, FL, July 1992; Central Florida Chapter of the Human Factors Society, Orlando, FL, July 1992; Florida Conference on Cognition, Gainesville, FL, April 1992; MLD Conference, Bloomington, IN, July 1992; Mid-Atlantic meeting of the Human Factors Society, Feb. 25, 1993, Chesapeake, Va.; Ohio State University, June 23, 1993, Columbus, Ohio; Loyola University, July 29, 1993, Chicago, Illinois.

Montgomery, D. A. and Sorkin, R. D. (1991). The effects of time stress on spatially selective operations. Paper presented at the 35th Annual Meeting of the Human Factors Society, San Francisco, CA, September, 1991.

Montgomery, D. A. and Sorkin, R. D. (1992). Observer sensitivity to element reliability in a multielement visual display. Paper presented at the 36th Annual Meeting of the Human Factors Society, Atlanta, GA, October, 1992.

Sadralodabai, T. and Sorkin, R. D., Discrimination of tonal sequences composed of repeated temporal patterns. Journal of the Acoustical Society of America, 1992, 91, 2333.

Sorkin, R. D., and Dai, H. Group Decision Making: Analysis of the Ideal Group. 33rd Meeting of the Psychonomics Society, St. Louis, MO, November, 1992.

Montgomery, D. A., & Sorkin, R. D. The effects of time stress on visual spatial selectivity. 35th Annual Meeting of the Human Factors Society, San Francisco, CA, 1991.

Montgomery, D. A. and Sorkin, R. D. Observer sensitivity to source reliability in a multi-element visual display. 36th Annual Meeting of the Human Factors Society, Atlanta, GA, 1992.

Sadralodabai, T., and Sorkin, R. D. Discrimination of tonal sequences composed of repeated sequences. Journal of the Acoustical Society of America, 1992, 91, 2333 (abstract).

Sadralodabai, T., and Sorkin, R. D. Serial position effects in temporal pattern discrimination. Journal of the Acoustical Society of America, 1993, 93, 2385-2386 (abstract).

Sorkin, R. D. and Dai, H. Psychoacoustic models of group signal detection. Journal of the Acoustical Society of America, 1993, 93, 2366 (abstract).

Sorkin, R. D. Models of Temporal Pattern Discrimination. ONR Conference on Neural Mechanisms of Temporal Pattern Perception, April 30, 1993, Durham, North Carolina; AFOSR Program Review, June 24, 1993, Dayton, Ohio.

5.2 Consultative and advisory functions.

Executive Board, Council of Graduate Departments of Psychology (COGDOP),

Member, 1992-1995; Secretary, 1993-1994; Chair, 1994-1995.

Acoustical Society of America Long Range Planning Committee, Chair, 1991-1994; Member, 1989-1994.

Member, Research Advisory Council, Center for Applied Human Factors in Aviation, Orlando, Florida, 1990-.

Faculty Advisor, University of Florida Chapter of the Human Factors Society, 1989-.

Member, National Research Council Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) 1987-1990.

Associate Editor, International Journal on Human-Computer Interaction. 1989-1993.

THE EFFECTS OF DISPLAY CODING FACTORS ON OBSERVER
VISUAL SIGNAL DETECTION

By

DEMARIS A. MONTGOMERY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

Abstract of Dissertation Presented to the Graduate
School of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

THE EFFECTS OF DISPLAY CODING FACTORS ON OBSERVER
VISUAL SIGNAL DETECTION

By

DeMaris A. Montgomery

August, 1993

Chairman: Robert D. Sorkin
Major Department: Psychology

Two studies examined the effects of display factors on

observers' ability to use multiple sources in visual signal

detection.   The information provided by a given source was

represented as a value on a graphical element.   Each dis-

played value was an independent sample from one of two

normal distributions, depending on the type of trial (Signal

or Noise) and the task being performed (Yes/No or Four-

Alternative-Forced-Choice, 4AFC).

The first study examined observers' ability to use

differences in source reliability in performing a Yes/No

decision task.   The reliability of the different display

elements was controlled by manipulating the variance of the

distributions from which the element values were sampled

(high reliability = low variance). Observers' efficiency in

weighting the sources based on their reliability was

estimated. Observers were relatively inefficient at using reliability information in forming a two-alternative decision (signal or noise). Only when a luminance cue to source reliability was introduced at stimulus durations equal to or greater than 400 ms was observer performance equivalent to an equal reliability condition. The evidence suggests that luminance cues aid observers in prioritizing visual information sources according to their importance to the task.

The second study examined the effects of display element arrangement on observers' performance in both Yes/No and 4AFC visual signal detection tasks. The information was displayed graphically in one of six formats constructed from a combination of two factors: 1) whether or not the display elements were arranged to produce a global feature that resulted from the interaction of the separate display elements, an "emergent feature," and 2) whether or not the magnitude of this global feature was monotonically related to the optimal decision statistic (for the Yes/No task). The results indicate that performance was facilitated by an emergent feature in the Yes/No task and was hindered by the presence of an emergent feature in the 4AFC task. Due to the relatively high performance produced by an angular element code, it was not possible to determine whether visual signal detection was affected by the presence of a relationship between an emergent feature and the optimal decision statistic.

MATHEMATICAL MODELS

Introduction

Every day, humans are faced with uncertain circumstances in which they have to form decisions based on multiple sources of information. In some situations these decisions have to be made rapidly, possibly to avoid an unfortunate outcome. For example, air traffic controllers have to detect and respond to selected events under time stress in order to avoid potential aircraft collisions. In many situations, the information is conveyed to the decision maker via visual displays. As a result, researchers are interested in determining how efficiently observers can combine spatially and temporally presented visual information sources, and in identifying the factors which influence overall processing efficiency.

The current investigation examines observers' use of multiple, spatially presented, independent, visual information sources in forming detection decisions. Using the Theory of Signal Detectability (TSD, Green & Swets, 1966) paradigm, we can specify the performance of an optimal observer in different detection tasks. The central theme of this investigation was to identify whether selected display coding factors, partly derived from knowledge of the optimal observer, would assist observers' detection decisions.

1

The first study examines observers' ability to combine nine independent, informational sources to form a Yes/No detection decision (Signal or Noise). The information is coded as graphical elements in the visual field, and in some conditions the sources differed `in their reliability. Frequently, decisions are based on multiple sources of information that differ in their informativeness (or reliability). An optimal observer includes this information in her detection decision. That is, she weights the information according to informativeness. However, when decisions need to be made rapidly, observers do not always consider all relevant information. The observer may not consider all sources or she may not apply an optimal weighing strategy. Thus, the main concern of this study was to determine whether selected factors assist observers in directing their attention to more reliable informational sources.

The second study further examines the effects of selected display formats on observers' detection decisions. Observers were given four informational sources to perform either a Yes/No task, as in the first study, or a Four-Alternative-Forced-Choice (4AFC) detection task. Bennett and Flach (1992) summarize the results from a number of studies which suggest that factors related to the display element arrangement can differentially affect performance in these two detection tasks. That is, selected display arrangements are more likely to facilitate performance in tasks which require integration of information, as in a

Yes/No task, than performance in tasks which require focused attention, as in a 4AFC task. This study attempts to identify the importance of two factors related to display element arrangement which may be contributing to possible differences in performance between the two tasks.

In a Yes/No decision task, an observer is given a sample of n independent elements $(x_1, x_2, \ldots x_n)$ to decide which of two alternative events (signal or noise) led to the evidence observed. On a given trial, one of the two stimulus alternatives is true, and each element conveys independent information about the current state. On signal trials, each x_i is drawn from a normal distribution with a mean of μ_s and a standard deviation of σ_s. On noise trials, each x_i is drawn from a normal distribution with a mean of μ_n and a standard deviation of σ_n. Alternatively, in a 4AFC task, on each trial four independent elements are presented. The values of three of the sources are drawn from the noise distribution and one source value is drawn from the signal distribution. The observer has to decide which source represents the "signal" event.

Employing the TSD paradigm, we can use the information about the underlying distribution parameters to identify how an optimal observer should perform in each of these tasks That is, we can identify the optimal performance level of an observer who is only limited by the uncertainty of the evidence, and who uses an optimal decision strategy. Given this information, we can 1) identify how well an observer

performs relative to the ideal, and 2) attempt to facilitate observer performance, which is generally inferior to the ideal, by presenting the information in a manner which helps them to act like a mathematically ideal observer.

Defining the Optimal Observer in Yes/No Detection

The Theory of Signal Detectability (TSD, Green & Swets, 1966; Green, 1992) provides a quantitative model for describing decisions based on uncertain evidence. Since it is a normative theory, it prescribes an optimum means of combining the information to form a statistic upon which an observer can base her decision. According to TSD, an optimal decision statistic is a likelihood ratio, or some value that is monotonically related to the likelihood ratio. A likelihood ratio is the ratio of the conditional probabilities for the current trial evidence, x. That is,

$$L(x) = f(x|s) \: / \: f(x|n). \tag{1}$$

It is assumed that the underlying distributions are normal such that the conditional probabilities can be expressed as

$$f(x|n) = [1/(2\pi\sigma_n^2)^{\frac{1}{2}}]EXP[-\tfrac{1}{2}((x - \mu_n)/\sigma_n)^2] \quad \text{and}$$

$$f(x|s) = [1/(2\pi\sigma_s^2)^{\frac{1}{2}}]EXP[-\tfrac{1}{2}((x - \mu_s)/\sigma_s)^2], \tag{2}$$

where $\sigma_s = \sigma_n = \sigma_e$, and to simplify the derivations $\mu_n < \mu_s$. For n independent sources of information, x_1, x_2, ..., x_n, by definition the probability of their joint occurrence is the product of their separate probabilities, $p(x_1 * x_2) = p(x_1)p(x_2)$. Similarly, the likelihood ratio for multiple independent sources can be expressed as the product of the separate likelihood ratios, $L(x_1 * x_2) = L(x_1) * L(x_2)$.

Since the natural logarithm of the likelihood ratio is monotonic with the likelihood ratio, the $\ln L(X)$ is also an optimal decision statistic. Thus, we have the following equation:

$$Z = \ln L(x_1, x_2, \ldots x_n) = \ln L(x_1) + \ln L(x_2) + \ldots + \ln L(x_n). \quad (3)$$

When the definitions of the conditional probabilities for the likelihood ratios are included in equation 3 and this equation is reduced, it turns out that the optimal decision statistic is a weighted sum of the evidence (see appendix A for the derivation),

$$Z = \sum_1^n x_i ((\mu_{si} - \mu_{ni})/\sigma_{ei}^2) - \tfrac{1}{2}((\mu_{si}^2 - \mu_{ni}^2)/\sigma_{ei}^2), \quad (4)$$

where x_i is the ith source, drawn from either a signal distribution, $\mathrm{Normal}[\mu_{si}, \sigma_{ei}]$, or the noise distribution, $\mathrm{Normal}[\mu_{ni}, \sigma_{ei}]$. Given a large sample, Z is also normally distributed since it is the sum of n mutually independent random variables, and its mean and variance given the two alternatives are

$$E(Z|s) = \sum_1^n \tfrac{1}{2}((\mu_{si} - \mu_{ni})/\sigma_{ei})^2, \quad E(Z|n) = -\sum_1^n \tfrac{1}{2}((\mu_{si} - \mu_{ni})/\sigma_{ei})^2 \quad (5)$$

$$VAR(Z) = \sum_1^n ((\mu_{si} - \mu_{ni})/\sigma_{ei})^2. \quad (6)$$

The assumption is that on a given trial an ideal observer will compare Z to some decision criterion, D. If Z is greater than or equal to D, then the observer should respond "signal." Otherwise she will say "noise."

When the separate pieces of evidence are not equally reliable, an ideal observer is sensitive to these differences and weighs the informational sources accordingly. The informativeness, and thus the appropriate weight, for a given source can be represented by the d' statistic as follows:

$$d'_i = (\mu_{si} - \mu_{ni}) / \sigma_{ei}. \qquad (7)$$

Expected optimal performance in a detection task is limited by the informativeness of the underlying evidence. That is, an observer's performance given a single source will not exceed d'_i. Based on equation 7, the informativeness of a particular source can be manipulated by either changing the distance between the two distribution means, $\mu_{si} - \mu_{ni} = \delta\mu_i$, or the size of the standard deviation, σ_{ei}.

In the first study, $\delta\mu_i$ is held constant $(\delta\mu_1 = \delta\mu_2 = \ldots \delta\mu_n)$, and the informativeness of the separate sources is controlled by changing σ_{ei}. Smaller values of σ_{ei} produce larger d'_i values and thus represent more informative sources. Table 1 lists the distribution parameters corresponding with a condition in which the even sources, x_i, have lower variability, making them relatively more informative.

The optimal weight for a given source, $\acute{a}_i$, is related to the d' value. When μ_{si} is held constant, $\acute{a}_i$ is proportional to the reciprocal of the variance for that source,

$$\acute{a}_i = 1 / [\sigma_{ei}^2 \sum_{1}^{n} (1/\sigma_{ei}^2)]. \qquad (8)$$

Table 1.
The mean and standard deviation of five informational sources in which the sources alternate in reliability; the even elements have the highest reliability.

element	1	2	3	4	5
signal					
μ_s	1	1	1	1	1
σ_s	1.5	0.75	1.5	0.75	1.5
noise					
μ_n	0	0	0	0	0
σ_n	1.5	0.75	1.5	0.75	1.5
d'	0.67	1.33	0.67	1.33	0.67

When an observer's decision is based on multiple sources, optimal performance is expressed in terms of the following d' statistic:

$$d'_z = \sum_1^n (\delta\mu_i^2/\sigma_{ei}^2) \; / \; (\sum_1^n (\delta\mu_i^2/\sigma_{ei}^2))^{\frac{1}{2}} = [\sum_1^n d'^2_i]^{\frac{1}{2}} \qquad (9)$$

Equation 9 can, then, be rewritten to include the optimal weights as follows:

$$d'_{ideal} = [\delta\mu_i \sum_1^n \acute{a}_i] \; / \; [(\sum_1^n \acute{a}_i^2 \sigma_{ei}^2)^{\frac{1}{2}}]. \qquad (10)$$

If the weights are normalized and the optimal weighting pattern requires equal weight across elements, then equation 10 equals the product of the square-root of n and d'_i.

The preceding equations allow us to define the performance of an ideal observer who is only limited by the uncertainty of the evidence. This provides a standard by which we can compare an observer's performance which is frequently inferior to the ideal. For a multiple observation task, some of the loss in performance may be a product of observers using a nonoptimal weighting strategy. Other loss may be more generalized (e.g., some form of internal noise), showing up as an overall performance loss. To discriminate the effects of these two sources of error, we need to identify how the observer weights the separate sources. A technique designed by Berg (1989, 1990) provides a means for estimating the observer's relative weights.

Generally speaking, the observers' weights are related to the slopes of empirical cumulative normal distributions

which Berg refers to as Conditional-On-A-Single-Stimulus, or COSS functions. A COSS function is a plot of the proportion of times an observer responded "signal" as a function of the magnitude of a given element across experimental trials. Two COSS functions are calculated for each element, one for signal trials and one for noise trials.

Figure 1 shows the COSS functions derived from simulated data of an observer using three informational sources to perform a Yes/No detection decision. The COSS functions on the left represent an observer using an equal weighting strategy. The COSS functions on the right represent an observer who weights the first source most and the third source the least. The upper curves with the square symbols in each graph of figure 1 represent the COSS functions for the signal trials. The lower curves with the circles in each graph represent the COSS functions for the noise trials. Figure 2 depicts the weights of the three sources derived from the COSS functions shown in Figure 1. The small squares and the circles represent the weight estimates for the signal and noise trials, respectively. The lines connecting the points represent the average of the two weight estimates. The solid and dashed lines represent the equal and unequal weighting strategies, respectively. By comparing figures 1 and 2, it can be seen that when the COSS functions have similar slopes the weights are relatively equal. Alternatively, looking at the graphs on the right side of figure 1, we can see that the smaller weights
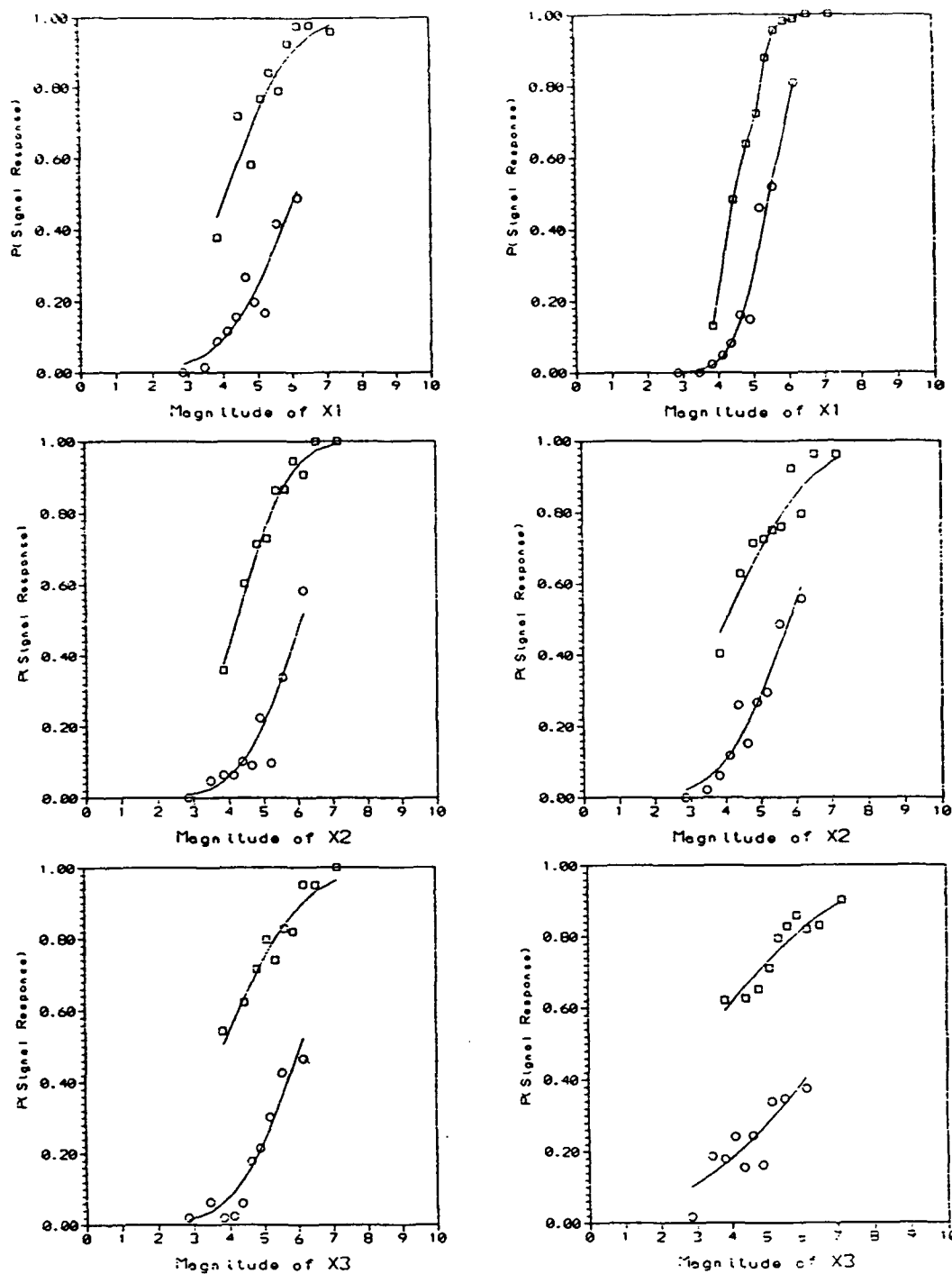
Figure 1.   The COSS functions derived from a simulated observer using an equal, left panels, or unequal, right panels, weighting strategy.   The top functions with the squares and the the bottom functions with the circles represent the signal and noise trials, respectively.
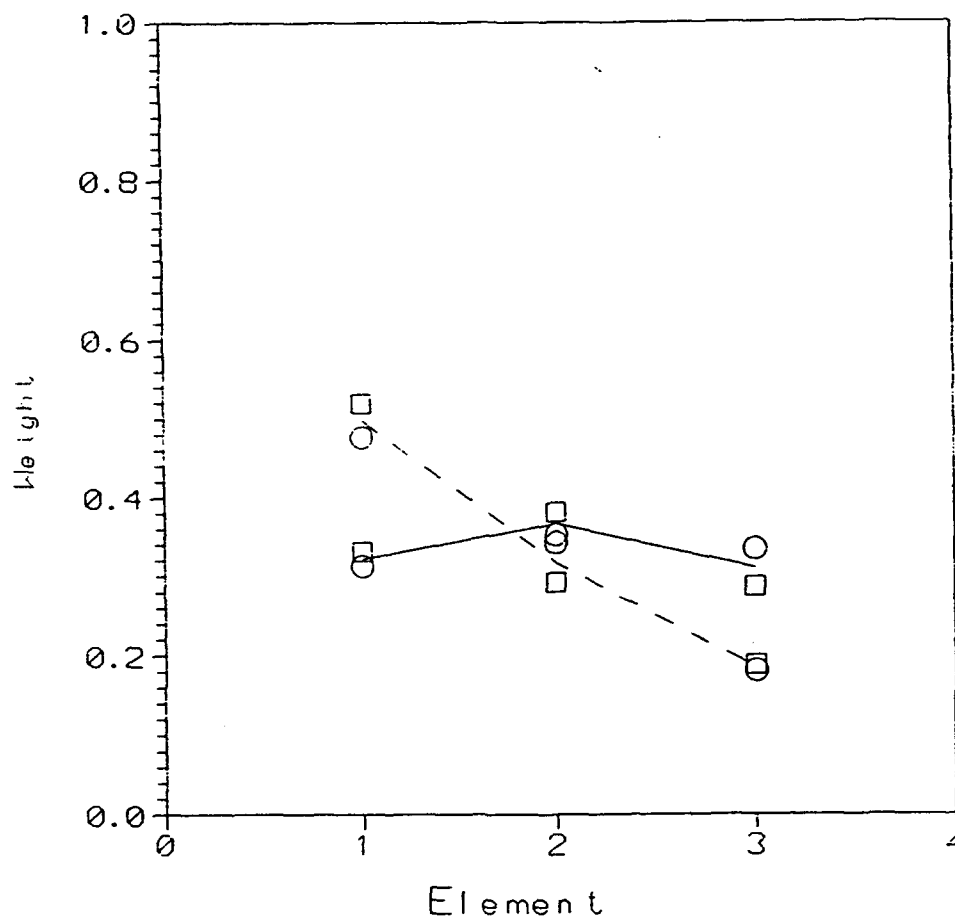
Figure 2. The weights derived from the COSS functions
depiected in Figure 1. The squares and circles
represent the weight estimates for the signal and noise
trials, respectively. The solid and the dashed lines
represent the average of signal and noise weights for
the equal and unequal weighting strategies, respect-
ively.

correspond with the shallower slopes in the COSS functions.

The actual weights depicted in Figure 2 are based on the variance of a cumulative normal distribution which had the best Chi-square fit, $VAR[Y_i]$, to the observer's COSS functions, represented by the solid lines in figure 1. (A detailed description of Berg's (1989) theoretical solution for the relative weights is found in appendix B). This estimated variance is added to the variance of the distribution from which the items were sampled, σ_{ei}^2. Then, to derive the relative weights, the sum of the variances for each source is divided by the sum of the variances corresponding with one source set to unity,

$$\frac{VAR[Y_j] + \sigma_{ej}^2}{VAR[Y_k] + \sigma_{ek}^2} = \frac{\sum_1^n a_i^2 \sigma_{ei}^2 / a_j^2}{\sum_1^n a_i^2 \sigma_{ei}^2 / a_k^2} = \frac{a_k^2}{a_j^2}. \tag{11}$$

Finally, the weights are normalized such that $\sum a_i = 1$.

Note that the choice of which source is to be set to unity is arbitrary. That is, the investigator should decide which item is the best choice given the hypothesis that is being addressed. For instance, Berg and Green (1990) used the COSS technique in an auditory profile analysis task. A profile task involves detecting an increment in the level of a single component (tone) among a multi-component background. Given an optimal decision strategy that compares the mean level of the signal component to the mean level of the nonsignal components, the greater the difference from

zero the more likely an increment was added. If the weight assigned to the signal component is set to unity, then the optimal weighting for the nonsignal components should equal $-1/(n-1)$ (where there are n components). For this experiment the element found in the center of the visual field at the fixation point, x_s, was set to unity.

Given the observers' estimated weights, Berg (1990) shows how these weights can be incorporated into a measure of the observers' weighting performance. This measure is the same as equation 10, except the observer's weights, a_i, are used instead of the ideal weights, $\dot{a}_i$,

$$d'_{wgt} = [\delta\mu_i \sum_1^n a_i] / [(\sum_1^n a_i^2 \sigma_{ei}^2)^{\frac{1}{2}}] \qquad (12)$$

If the observer applies a nonoptimal weighting pattern the observer's weighting performance, d'_{wgt}, will be lower than that of an ideal observer, d'_{ideal}.

Furthermore, we can obtain a measure of the observer's overall performance, d'_{obs}, on the task by calculating the absolute value of the difference between the Z-scores corresponding with her hit and false alarm probabilities on the experimental trials. If this measure, d'_{obs}, is lower than d'_{wgt}, then the additional loss in performance can be thought of as the effects of the observer's internal noise, σ_{int}. That is, unlike an ideal decision maker, an observer will often be less reliable at transferring information from the environment into a decision statistic. It is assumed that internal noise is independent of the weight estimates.

Finally, once the three performance measures, d'_{ideal}, d'_{wgt}, and d'_{obs}, have been derived, performance can be summarized in terms of an efficiency measure (Tanner & Birdsall, 1958). Berg (1990) describes observers' performance in terms of three efficiency measures: one representing the observers overall performance, another representing the observers' weighting performance, and a third representing residual factors such as internal noise. A general measure of the observer's overall efficiency, η_{obs}, is the squared ratio of her performance, d'_{obs}, relative to the performance of an ideal observer, d'_{ideal}. That is,

$$\eta_{obs} = (d'_{obs} / d'_{ideal})^2 . \qquad (13)$$

If the observer is optimal, $\eta_{obs} = 1.0$. Any decrement in the observer's performance will correspond with a decrease in efficiency, where $0 < \eta_{obs} < 1$.

The other two efficiency measures, η_{wgt} and η_{noise}, allow us to separate the loss in observer efficiency due to non-optimal weighting from loss due to observer internal noise, respectively. The weighting efficiency, like the overall efficiency, η_{obs}, is the measure of the observer's weighting performance, d'_{wgt}, relative to the ideal observer, d'_{ideal}, who uses an optimal weighting strategy:

$$\eta_{wgt} = (d'_{wgt} / d'_{ideal})^2 . \qquad (14)$$

η_{noise} accounts for any additional loss in d'_{obs} not explained by the weights,

$$\eta_{noise} = (d'_{obs} / d'_{wgt})^2 . \qquad (15)$$

The relationship among these measures is

$$\eta_{obs} = \eta_{wgt} * \eta_{noise}.$$

(16)

Defining the Optimal Observer in 4AFC Detection

In a Four-Alternative-Forced-Choice (4AFC) task, an observer is given four independent sources of information, where each source represents one of two alternatives, "signal" or "noise." On a given trial, one of the four elements is randomly selected to represent the signal event. This source value is drawn from a normal distribution with a mean of μ_s and a standard deviation of σ_s. The remaining three source values are drawn from a normal distribution with a mean of μ_n and a standard deviation of σ_n, where $\mu_n < \mu_s$ and $\sigma_n = \sigma_s = \sigma_e$. The observer's task is to identify which of the four sources represents the signal. Rather than combining the information to make a single response, as with the Yes/No task, a 4AFC task requires the observer to independently assess each value to identify which source represents the signal.

In the second study, there are four informational sources, represented by graphical elements located in four separate spatial position in the visual field. For the 4AFC task there are four possible stimulus sequences, <s,n,n,n>, <n,s,n,n>, <n,n,s,n>, or <n,n,n,s>, where <s,n,n,n> represents a stimulus value in the first spatial position and three noise values in the second, third, and fourth spatial positions. The observer has to determine which of the four spatial orders in present on a given trial. Thus, there are

four possible responses, $<S,N,N,N>$, $<N,S,N,N>$, $<N,N,S,N>$, or $<N,N,N,S>$, corresponding with the four equally likely locations where the signal can occur.

Table 2 depicts the stimulus-response matrix for the decision task. The matrix cells falling along the minor diagonal represent correct responses. T_i represents the total correct responses for the ith ordering of the stimuli, Sg_i. The percentage correct in a 4AFC task, $P_4(C)$, is

$$P_4(C) = \sum_1^n T_i / N_{Total}, \qquad (17)$$

where N_{Total} is the total number of trials across all stimulus orders.

Green (1992) shows that an ideal observer, who attempts to maximize percent correct, will choose the source with the largest value since this value also has the largest likelihood ratio. To expedite the derivation, Green characterizes the task as detection of 1-of-m possible signals relative to noise alone. Using this approach, each sequence would be represented as a separate signal, e.g., $<s,n,n,n> = Sg_1$. Thus, the likelihood that the evidence, $x = <x_1,x_2,x_3,x_4>$, presented on a given trial represents the ith signal compared to noise alone may be expressed as follows:

$$l(x|Sg_i) = EXP[x_i((\mu_s-\mu_n)/\sigma_e^2) - \tfrac{1}{2}((\mu_s^2-\mu_n^2)/\sigma_e^2)], \qquad (18)$$

where $\mu_n < \mu_s$, and $\sigma_n = \sigma_s = \sigma_e$. x_i is monotonically related to the optimal decision statistic, $l(x|Sg_i)$. Thus, the observer should choose the largest value, x_i, since this value also has the largest likelihood ratio. Green's (1992)

Table 2.

The stimulus-Response matrix for the 4AFC task. The sequence <s,n,n,n> represents a stimulus value in the first spatial position and three noise values in the second, third, and fourth spatial postitions.

	Sg1 <s,n,n,n>	Sg2 <n,s,n,n>	Sg3 <n,n,s,n>	Sg4 <n,n,n,s>
<S,N,N,N> T1				
<N,S,N,N>	T2			
<N,N,S,N>		T3		
<N,N,N,S>			T4	

derivations of the optimal decision statistic for an Four-Alternative-Forced-Choice task are found in appendix C. An alternative calculation of the optimal decision statistic that considers the sequences as four separate hypotheses yields a slightly different decision statistic, but the same decision strategy. That is, an optimal observer should choose the source with the largest value since it also has the largest likelihood ratio.

The accuracy of an observer using this predicted optimal decision strategy depends upon the probability that the largest value was actually sampled from the signal distribution. That is, $P_4(C)$ depends on the probability that the sample from the signal distribution, $f(x|s)$, is greater than the samples from the noise distribution, $f(x|n)$. Considering two alternatives, the probability that one random variable is larger than another can be expressed as follows:

$$P_2(C) = \int_{-\infty}^{\infty} f(u|s) \int_{-\infty}^{u} f(v|n) \, dv du. \tag{19}$$

Equation 19 represents the probability that the value of the noise sample, v, is less than the value of the signal sample, u, summed across all possible values of u (Green, 1992). Since the same probability density functions, $f(x|s)$ and $f(x|n)$, are used to define the hit and false alarm probabilities found in a Yes/No ROC curve, it is possible to relate performance in an m-Alternative-Forced-Choice task to performance in a Yes/No task. That is, equation 19 can be rewritten to produce the following equation,

$$P_2(C) = \int_0^1 [1-P_u(S|n)]\,dP_u(S|s), \qquad (20)$$

where $1-P_u(S|n)$ is the complement of the false alarm probability, and $-dP_u(S|s)$ is the derivative of the complement of the hit probability. (See Appendix D for the derivations.) Equation 20 shows the area under a Yes/No ROC curve is related to percent correct in a 2AFC task (Green, 1992; Green & Swets, 1966). Finally, equation 20 can be rewritten to account for multiple alternatives as follows,

$$P_m(C) = \int_0^1 [1-P_u(S|n)]^{m-1}\,dP_u(S|s), \qquad (21)$$

where $m > 2$.

Thus, it is possible to convert a percent correct value in an m-Alternative-Forced-Choice task to a Yes/No d' measure. Hacker and Radcliff (1979) published tables which allow us to make conversions from percent correct in an m-Alternative-Forced-Choice task to a Yes/No d'. This table takes into account the uncertainty associated with larger numbers of alternatives. For instance, when $P_2(C) = 0.8$ in a 2AFC task d' = 1.19; however, in a 4AFC task the same percent correct, $P_4(C) = 0.8$, yields a d' = 1.89. Finally, given the relationship between performance in Yes/No and mAFC tasks, maximum percent correct in a 4AFC task will also be limited by the underlying distribution parameters (MacMillian & Creelman, 1991). That is, when percent correct in a 4AFC task is converted to a Yes/No d', performance will not exceed d' as defined in equation 7.

The preceding definitions of the optimal observer in both a Yes/No and 4AFC detection decision, provide a baseline for comparing observer performance under different experimental conditions. In situations where observer performance falls short of the ideal, performance may be facilitated by presenting the information in some manner which helps them to act more like an ideal observer. The two studies, to be described, address this approach to optimizing human performance. That is, these studies look at the effects of selected display coding factors which were designed to help observers function as optimal observers on their detection decisions.

EXPERIMENT 1

Introduction

Visual displays are commonly used to convey system information, such as air traffic flow or the status of a production line, to a human decision maker. A complex visual display may include several subordinate displays or display "elements." Each display element provides a potential source of information for the human operator. However, it may be impossible for the operator to obtain useful information from more than a few of the display elements at one time. This problem may be minimized if the operator can prioritize the display elements in terms of their criticality and informativeness, and if the operator can allocate his or her attention accordingly. This study examined several factors that affect an operator's ability to allocate attention to display elements that are differentially informative.

In a previous experiment (Sorkin, Mabry, Weldon, & Elvers, 1991), observers examined a multi-element display and then reported whether the display represented the occurrence of a signal or nonsignal event. Using a technique derived from the Theory of Signal Detectability (TSD, Green & Swets, 1964), Sorkin et al. estimated the importance or weight the observer assigned to each element of the display

21

in making a detection decision (Berg, 1989, 1990). An optimal decision-theoretic observer weights the input from each element according to the element's informativeness or reliability; highly reliable display elements are weighted more highly in the detection decision than less reliable elements (Durlach, Braida, & Ito, 1986; Berg, 1990; Berg & Green, 1990).

In the Sorkin et al. (1991) study all display elements were equally informative; hence, each element should have been weighed equally in the observers' decisions. When the observation durations were long, the weights were equal across the spatial array of display elements. However, when the observation durations were brief and the display coding was complex, the highest decision weights were associated with display elements in the center of the visual field, around the observer's fixation point. The extent to which the weighting functions were peaked corresponded with the performance level (low performance was associated with peaked functions). Sorkin et al. (1991) concluded from these results, that under difficult conditions, the observer's allocation of attention was restricted to the central portion of the display.

This interaction between the difficulty of the task and the availability of information from different regions of the display is not surprising. A number of variables are known to affect an observer's ability to obtain information from the elements of a complex visual display. These

include the number (Perrott et al.; 1991) and spacing (Andre & Wickens, 1988) of irrelevant, or distracter, items found in the visual field, the type of display code (Boles & Wickens, 1987; Legge, Gu, & Luebker, 1989; Sanderson, Flach, Buttigieg, & Casey, 1989; Sorkin et. al., 1991), display item intensity (Eriksen & Rohrbaugh, 1970), and task complexity (Williams, 1982).

When the stimulus durations in the Sorkin et al. (1991) experiment were long (more than 500 ms), all display element weights were equal, indicating that the observers could process information from all regions of the display. Since the reliability of all the elements was also equal, an equal weight strategy was optimal for that task. An important question is whether an observer can employ optimum weights when the reliabilities of the elements are not equal across the visual array. Obviously, the ability to match decision weights to the element reliability is necessary if the observer is to prioritize the display elements according to their importance to the task.

When an informational source does not provide a consistent report of an unchanging event, the source is not reliable. For instance, if a sensor measures a specific luminance value to be x at one time and x ± n on a subsequent reading, the sensor is showing variability in its measurement. Thus, this sensor would be less reliable than one which produces a consistent measure across time. A person forming a decision based on this information should place

greater weight on the more reliable source. However, evidence suggests that people tend to overrate the importance of unreliable sources (Schum, 1975). Wickens (1984) states that when people are confronted with sources which are not equally informative, they perform the task "as if" all sources were equally reliable.

The present study addressed whether observers can use differences in display element variability to identify source reliability and use this information in forming a simple Yes/No detection decision. In addition, this study was designed to determine whether using this information imposes a significant amount of additional processing "overhead" on the observer, and whether selected display factors could reduce related performance loss. As in Sorkin et al. (1991), the observers in the current study performed a multi-channel visual detection task. On each trial of the experiment, observers were presented with a display consisting of nine display elements. The display elements were nine vertical line-graph gauges arranged in a horizontal array (see figure 3). The values displayed on the line-graph gauges, $<x_1, x_2, \ldots, x_9>$, were determined by independent, normally distributed, random variables. On a signal trial, the values of the nine elements were selected from a distribution with a mean of μ_s and a standard deviation of σ_s. On a noise trial, the values were drawn from a distribution with a mean of μ_n and a standard deviation of σ_n, where $\mu_n < \mu_s$. The observer's task was to decide
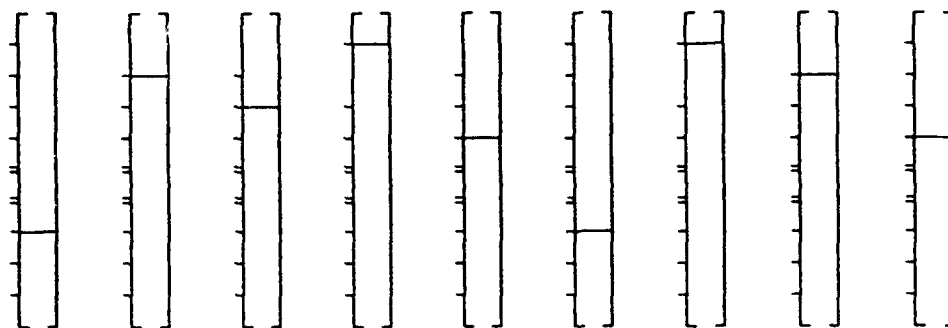
Figure 3. Demonstration of the nine graphical elements found in experiment 1.

whether the data displayed had been generated from the signal or noise distribution.

The reliability of different display elements was controlled by manipulating the variance of the distributions from which the element values were sampled: high reliability elements were sampled from distributions with lower variance than low reliability elements. That is, a source high in reliability would be analogous to an instrument which shows consistent measurements across time whereas a source low in reliability would not provide consistent measurements. The variance of a display element at a particular position was the same for signal and noise trials, but differed across elements depending on the experimental condition. Table 1 illustrates the mean and standard deviations that could be employed for a five element display in which odd and even elements alternated in their level of reliability.

Berg (1990) found that the reliability of elements in an auditory task similar to the one used in this study could be used by observers when the most reliable tones were much louder than the less reliable tones. The loudness cue was much less effective when reversed, in which case a louder cue indicated a lower reliability. Berg's results suggest that under some conditions cuing element reliability (e.g., with intensity or color) may aid observers in accurately weighting display sources by their importance.

Cues such as size, intensity, color, and movement are often incorporated in display design to draw attention to

specific items in a display. For instance, researchers have found that correct utilization of color coding (Christ, 1990; Fisher & Tan, 1989) can reduce search time in locating an item in a display. Furthermore, Wickens and Andre (1990) showed that color coding a particular item in an object display leads to improved accuracy in recalling the specific value associated with that item relative to a monochromatic display. Thus, given Berg's results and the evidence cited above, we predicted that observer weighting efficiency in the present experiment should be higher for a condition in which a luminance cue signals the element reliability.

In order to test the efficacy of a cue for element reliability in the present experiment, the spatial position of the high reliability display elements was randomly varied over trials. The overall luminance of the display element varied in accordance with the reliability (high or low) of the element. We expected that luminance would provide a natural code for allocating observer attention and hence weight, to the high reliability elements. If that were the case, the efficiency of the observers' weighting strategy would be much higher in a cued than in an uncued condition.

The duration of the stimulus and the spatial arrangement of the element reliabilities also should influence how efficiently the observers match their weights to the element reliabilities. The results from the Sorkin et al. (1991) study suggested that 233-ms was sufficient time for observers to utilize information from as many as nine, equally

reliable, graphically coded display elements. However, it is possible that sensing the element reliabilities and differentially weighting the elements, may require some additional processing steps or "overhead" by an observer. A duration of 233-ms may be at the margin of an observer's ability to extract the information needed to discriminate and employ differences in element reliability. For example, a slower, serial search may be required to extract the reliability information and weight the elements accordingly. In that case, it might be advantageous, when processing short duration stimuli, to ignore reliability and differential weighting information. Our experiments tested three levels of stimulus duration (150, 400 and 800-ms). We expected that weighting efficiency would be greatest at long stimulus durations (400-ms and 800-ms) and very poor at the shortest duration (150-ms).

Observer sensitivity to element reliability also may be affected by the spatial arrangement of element reliability. If attention is distributed more effectively among spatially contiguous than separated items, grouping sources similar in reliability should aid performance. Posner, Snyder, and Davidson (1980) found that simple reaction time to detect a light at a second most likely position was facilitated only when this item was adjacent to a cued location (the most likely target location). When the second most likely position was separated by more than one location, detection speed was not facilitated. Thus, the weighting efficiency

should be better for displays with elements grouped by similar reliabilities than for displays that distribute element reliabilities across the array.

Finally, we were interested in whether observers would be sensitive to the reliability of individual elements without any cues to element reliability. That is, can observers estimate (and employ) information about element reliability based only on the trial-by-trial variability of the readings from individual display elements and feedback about the S/N events? To answer that question, we added conditions in which the relationship between element spatial position and reliability was fixed, rather than random, over a block of 200 trials. If the observer can estimate the variance of the element readings from the first k trials of a block, the observer may be able to partition the elements into those with high and low reliabilities. If that process led to the assignment of higher weights to the more reliable elements, the observer's weighting efficiency would be enhanced in that condition.

Method

Subjects

Four University of Florida students with normal, or corrected to normal, visual acuity participated in this study. One subject, S2, was later discovered to be color deficient, and another, S4, was highly trained on the task. They were paid an hourly wage plus a bonus based on performance.

Apparatus and Stimuli

Observers were seated in a sound isolated booth approximately 27 inches away from a 10.5 inch color monitor (EGA) driven by an 80386 computer. The monitor was set for maximum contrast, and intensity was set at approximately 102 cd/m^2, measured from a 7.5 inch by 4 inch uniform white field. On a given trial, nine gauges were presented on the monitor; subtending a horizontal by vertical visual angle of approximately 16° by 8°. Each gauge was composed of two parallel white lines, with tick-marks falling at equal intervals on the left line for all conditions except the luminance cue condition. For this latter condition high reliability gauges were white and the remaining gauges were gray. The intensity of the white gauges was approximately 102 cd/m^2 and the intensity of the gray gauges was approximately 22 cd/m^2 measured from 7.5 inch by 4 inch uniform white and gray fields, respectively.

Each tick-mark represented a display increment of 1.0, and ranged from 0.0 to 10.0. Two longer blue lines, located near the tick-marks, indicated the positions of the signal and noise distribution means. The value displayed by each gauge was determined by sampling a number from either a "signal" or "noise" distribution, depending on the type of trial. This number was converted to the vertical displacement of a horizontal white line from the bottom (e.g. zero position) of the gauge (see figure 3). The gauge values were drawn from the signal distribution on 50 percent of the

trials. The mean of the gauge values on signal trials, μ_s, was equal to 5.0; the mean on noise trials, μ_n, was equal to 4.0.

The standard deviation of the gauge values on signal and noise trials depended on the particular experimental condition. There were five different element reliability conditions: (1) Equal, (2) Grouped-Left-High, (3) Grouped-Right-High, (4) Distributed-Even-High, and (5) Distributed-Odd-High. In the Equal condition, the standard deviation of all gauge elements was equal to 1. In the Grouped-Left-High condition, the standard deviation of the four left elements was equal to 0.75, and the five right elements was equal to 1.5. That pattern was reversed in the Grouped-Right-High condition. In the Distributed-Even-High condition, the standard deviation of the four even elements (element 2, 4, 6, and 8) was equal to 0.75, and the standard deviation of the remaining elements was equal to 1.5. In the Distributed-Odd-High condition, the standard deviation of the five odd elements (element 1, 3, 5, 7, and 9) was equal to 0.85, and the standard deviation of the remaining elements was equal to 1.3.

The unequal reliability conditions were run under two different trial block conditions: Pure Block and Mixed Block. In the Pure Block condition, all display and distribution parameters were fixed within a block of 200 trials. Thus, in four conditions (Grouped-Left-High, Grouped-Right-High, Distributed-Even-High, and Distributed-Odd-High), the

relationship between element reliability and spatial position was fixed throughout the block of trials. In the Mixed Block conditions, the trials within a block of 200 trials alternated randomly among the Grouped-Left-High, Grouped-Right-High, Distributed-Even-High, and Distributed-Odd-High conditions. In the Mixed Block conditions, it would be impossible for an observer to identify the reliability of any given spatial element, unless the observers were provided with an additional trial-by-trial cue to element reliability. Finally, all trial block conditions were tested at three levels of stimulus duration (150, 400, and 800 ms).

The duration of the stimulus presentation was synchronized with the refresh traces of the monitor. The period between traces was approximately 17 ms. The onset and offset of the display was delayed until a retrace was ready to occur. Once the stimulus was presented the duration was controlled by counting the number of refresh traces which corresponded with the selected stimulus duration (150, 400, or 800 ms).

The experimental conditions are shown in table 3. The mnemonics in each table cell describe the trial-block conditions. The three trial-block conditions which contained elements which differed in reliability across spatial positions are denoted by the letter U in the mnemonic, meaning the sources were unequal in reliability. The equal reliability condition is denoted by the letter E in the mnemonic. As demonstrated in the table all trial-block conditions

Table 3.
The mnemonics for experimental conditions found in
experiment 1.

		UNCUED			CUED		
		150	400	800	150	400	800
MIXED BLOCK	LEFT	UNcM			UCM		
	RIGHT						
	ODD						
	EVEN						
PURE BLOCK	LEFT	UNcP					
	RIGHT						
	ODD						
	EVEN						
EQUAL RELIABILITY		ENcP					

were run at the three levels of stimulus duration. In addition, within each of the unequal reliability trial-block conditions the four element reliability arrangements were presented.

The next two letters in the mnemonics represent whether or not a luminance cue was present (C = cue and Nc = No cue). Finally, the last letter, P or M, denotes the manner in which the element reliability arrangements were presented. The P represents a pure block design in which the arrangements remained constant across experimental trials in a given block, and the M represents a mixed block design in which the arrangements varied across trials. Thus, the mnemonic UNcM stands for an Unequal reliability No cue Mixed block design.

Procedure

Observers were told to make their decisions based on the level of the gauges relative to the signal and noise mean markers. They were told to rank the likelihood that the evidence represented a signal by using the "4", "3", "2" and "1" keys, where "4" represented very sure it was a signal and "1" represented noise. In fact, observers tended only to use the two middle keys. Thus, responses on keys "1" and "2" were combined to represent noise, and responses on keys "3" and "4" responses were combined to represent signal in the data analyses. On conditions where the reliabilities differed across elements, observers were informed that the least variable gauges were the most reliable.

The trial sequence, shown in figure 4, proceeded as follows. First, observers were given a 0.5" by 0.5" fixation cross at the center of the display for 200 ms. This was replaced by the nine line-graph gauges for a stimulus duration of either 150, 400, or 800 ms. Following the stimulus a white blanking mask was presented for 200 ms. Then, the display was completely black for 1 second, at which time the observers were allowed to respond. Any responses made prior to or following this period were discarded as "No Response" trials. Finally, the observers were given feedback at the center of the display for 250 ms. Within a given session, an observer ran through 10 blocks of 200 trials. Across sessions there were 1500 trials (750 signal and 750 noise) collected for each condition.

Due to time constraints imposed by the need to collect multiple trials, some of the observers received less practice than others. Subject S4 was highly practiced. He ran through at least eight practice sessions for each condition prior to collection of the experimental trials. Subjects S1, S2, and S3 were highly practiced on the Yes/No detection task, but they only ran through one practice session for each of the individual conditions.

To control for any possible practice effects in the experimental sessions, each observer received a different order of four trial block conditions, organized such that across subjects each condition occurred once in each the four possible positions in the order.

Fixation: 200 ms	+
Stimulus Duration: 150,400 or 800 ms	H H H H H
Mask: 200 ms	White Field
Response Duration: 1000 ms	Black Field
Feedback: 250 ms	Correct

Figure 4. Trial sequence for the first experiment.

Results

Average observer performance measures (d'_{obs}) for the experimental conditions are shown in table 4. In order to consider differences between the equal and unequal reliability block-type conditions the data were collapsed across source reliability arrangements for the unequal reliability conditions. An analysis of variance performed on the average d'_{obs} showed a significant main effect of stimulus duration $(\underline{F}(2,6)=12.49, \underline{p} \leq 0.01)$. Performance improved as stimulus duration increased. There was also a main effect of block-type condition $(\underline{F}(3,9)=5.285, \underline{p} \leq 0.05)$. All four observers showed greater performance in the equal reliability condition relative to the two unequal reliability conditions which did not include a luminance cue to the more reliable sources (UNcM and UNcP).

An analysis performed on the observers d'_{obs} measures for the unequal reliability conditions indicated significant effects for all of the experimental variables (block-type condition, stimulus duration, and arrangement), and their interactions, except for the three-way interaction. However, only a few of these differences were evident in the data of the individual subjects. All observers showed a performance improvement as stimulus duration increased $(F(3,6)=13.66, p\leq0.01)$. There was also a performance advantage for the cued block-type condition, UCM, relative to the UNcM and UNcP conditions $(\underline{F}(3,9)=3.86, p\leq0.05)$.

Table 4.
Average observer performance (d') for each experimental condition.

		UNCUED			CUED		
		150	400	800	150	400	800
MIXED BLOCK	LEFT	1.92	2.11	2.12	2.07	2.37	2.40
	RIGHT	1.77	1.89	2.08	1.90	2.21	2.34
	ODD	2.06	2.22	2.28	1.92	2.06	2.14
	EVEN	1.93	2.02	2.18	2.01	2.20	2.23
PURE BLOCK	LEFT	1.77	2.05	2.14			
	RIGHT	1.68	1.92	2.04			
	ODD	1.94	2.07	2.06			
	EVEN	1.74	1.98	2.03			
EQUAL RELIABILITY		2.24	2.44	2.57			

x 4 OBSERVERS

Finally, there was consistency among the observers for two of the interactions, as well. For the four arrangements there was a tendency for observer performance to be highest when the most reliable sources were grouped in the four left spatial positions at stimulus durations of 400 ms or greater. Alternatively, at the shortest stimulus duration performance was highest when the most reliable sources were distributed among the odd spatial positions. The effects of arrangement also depended on the particular condition. In general, observers showed a performance advantage for the Odd arrangement over the other three arrangements in the two, no luminance cue conditions (UNcM and UNcP). Alternatively, the grouped left arrangement yielded the greatest observer performance in the mixed block, cued condition (UCM). Moreover, the right arrangement tended to show the poorest performance in the UNcM and UNcP conditions, but relatively high performance in the UCM condition. All of these differences were found to be significant through subsequent paired comparisons using a Tukey test.

Thus, the evidence from these analyses indicated that stimulus duration and block-type condition had a consistent effect on observer's performance. In addition, the arrangement of source reliabilities influenced observers' performance. However, the direction of effects depended on the stimulus duration and the block-type condition. Performance was greatest when the stimulus duration was at least 400 ms and sources were equal in reliability. Considering the

unequal reliability patterns alone, performance was best when sources high in reliability were cued and grouped. However, performance was also relatively high for the odd arrangement in the UNcM condition. Since the location of reliable elements in the visual field affected performance differently under specific conditions, it is logical to suspect that the differences in performance are related to observers' weighting strategies. For example, when observers are under time constraints or there is uncertainty about the location of reliable sources, observers may be less efficient at applying weights appropriate to the weighting strategy selected.

Weight Analysis

Observers' weights were estimated using Berg's (1989, 1990) Conditional-On-A-Single-Stimulus (COSS) analysis technique, described earlier. The estimated weights were based on the slopes of cumulative normal functions that had the best Chi-Square fit with the corresponding COSS functions. Two weight estimates were calculated for each element in each condition, one for signal and one for noise trials. In this analysis, out of 2808 COSS functions Chi-square fitted to cumulative normals, 6.7% significantly differed ($p <= 0.05$) from the observers' COSS functions. The weights reported are the average of the signal and noise weight estimates.

Table 5 lists the mean weighting efficiencies derived from the weight estimates of the conditions found in the

Table 5.
Observer Weighting Efficiency Estimates for Stimulus
Duration and Condition.

| Subject | Condition | Stimulus Duration | | | |
		150	400	800	Average
S1:	ENcP	0.650	0.780	0.850	0.760
	UNcM	0.540	0.645	0.683	0.622
	UNcP	0.675	0.755	0.797	0.743
	UCM	0.735	0.835	0.850	0.807
S2:	ENcP	0.720	0.820	0.790	0.777
	UNcM	0.485	0.563	0.633	0.560
	UNcP	0.510	0.617	0.627	0.585
	UCM	0.607	0.740	0.795	0.714
S3:	ENcP	0.710	0.830	0.860	0.800
	UNcM	0.657	0.688	0.705	0.683
	UNcP	0.495	0.650	0.680	0.608
	UCM	0.715	0.758	0.802	0.758
S4:	ENcP	0.940	0.960	0.970	0.957
	UNcM	0.792	0.745	0.780	0.772
	UNcP	0.727	0.770	0.740	0.746
	UCM	0.900	0.958	0.960	0.939
Avg:	ENcP	0.750	0.848	0.868	0.823
	UNcM	0.619	0.660	0.700	0.660
	UNcP	0.602	0.698	0.711	0.670
	UCM	0.739	0.822	0.852	0.805

Equal reliabilty No cue Pure block design (ENcP)
Unequal reliability No cue Mixed block design (UNcM)
Unequal reliability No cue Pure block design (UNcP)
Unequal reliability Cue Mixed block design (UCM)

first ANOVA described earlier. Again, for the three unequal reliability conditions the efficiencies represent the average of the four arrangements. Across all observers and conditions, weighting efficiency increased as the stimulus duration increased. To identify whether these differences were significant, a Monte Carlo simulation was run to estimate the expected variance for η_{wgt}. The sampling distribution of η_{wgt} which best reflected the observers weighting efficiencies was used to identify significant differences.

The criterion selected for significant differences was two standard deviations ($\sigma_{sim} = 0.04$) of this sampling distribution. Differences in weighting efficiency which exceeded 0.08 were identified as significant differences. This was a fairly conservative estimate, considering that these distribution parameters corresponded best with the data from the poorest observer. Given this criterion, weighting efficiency was significantly greater for stimulus durations of 400 ms or higher, and efficiency was highest at 800 ms. Thus, significant improvements in weighting efficiency at least partially account for the improvement in overall accuracy found when stimulus duration increased.

Moreover, differences among the observers' weighting efficiencies for the block-type conditions are consistent with the differences observed in the overall d'_{obs} measures. Performance was greatest in the ENcP and UCM conditions. Both show a significant advantage over the other two unequal reliability conditions (UNcM and UNcP). This pattern is

maintained at all stimulus durations and fairly consistent across the four observers.

Since performance was highest at the longest stimulus duration, and differences among the conditions were consistent across all levels of stimulus duration, the figures depicting the separate observers' weights for the unequal reliability conditions are based on the data obtained at the 800 ms stimulus duration, only. Figures 5-16 show the observers' average weights for the four arrangements of source reliability and the three unequal reliability block-type conditions. There are three figures representing the three block-type conditions (UNcM, UNcP, and UCM) for each observer. In each figure there are four graphs representing the four source reliability arrangements, where a, b, c, and d represent the left, right, even, and odd arrangements, respectively. The larger symbols are the average weight estimates and the smaller symbols are the signal and noise weight estimates. The solid line represents the optimal weights for the separate arrangements. All four observers show similar changes in their weights across the three conditions, where their weights best match the ideal weights for the UCM condition.

Table 6 lists the observers' weighting efficiencies for the conditions found in figures 5-16. Again, only the 800 ms duration is shown. Table 7 shows the average weighting efficiencies for stimulus duration, arrangement, and block-type condition. As with the average data, the UCM condition
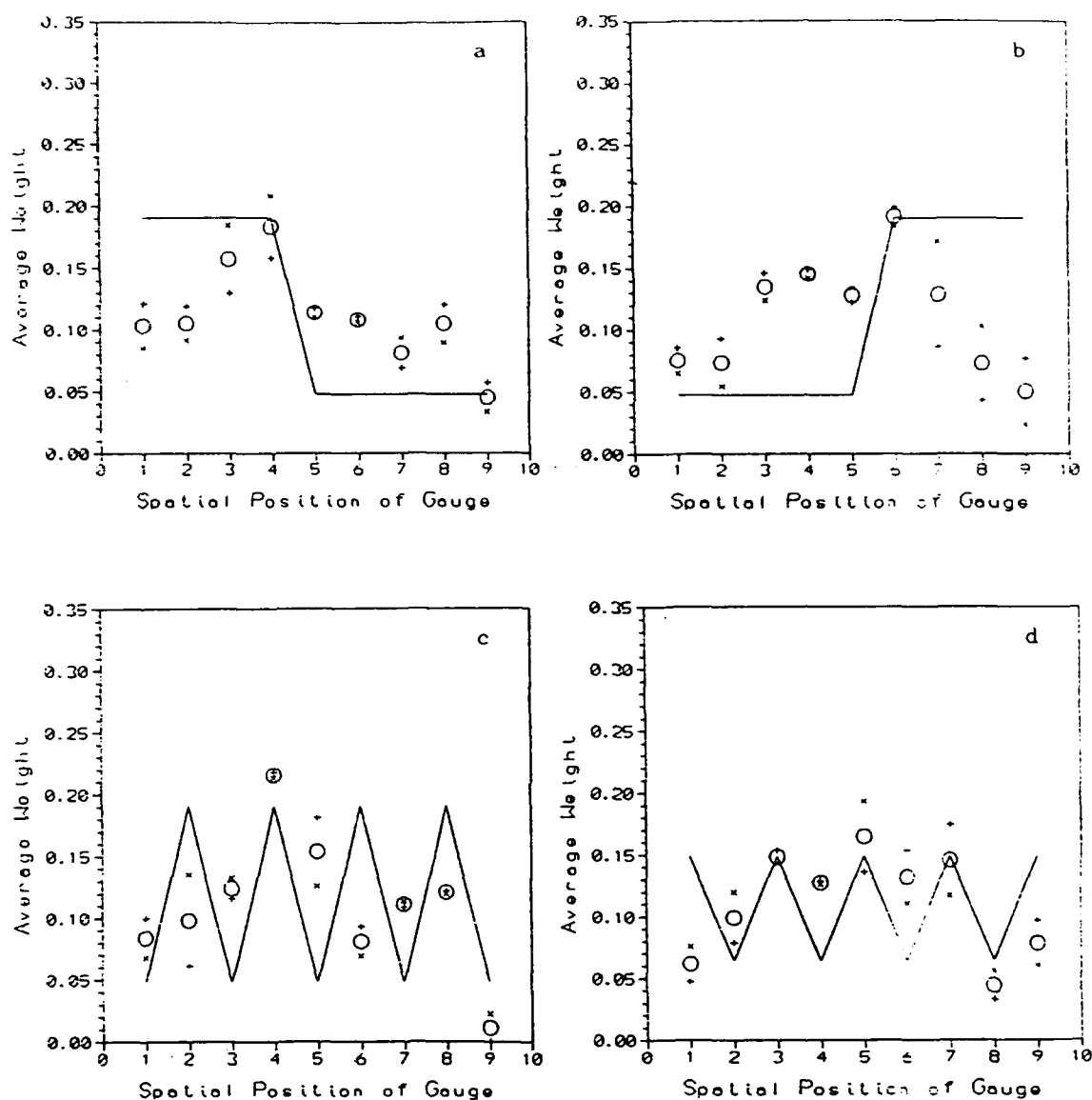
Figure 5. Subject S1's average weights for the four source reliability arrangements in the UNcM block-type condition. Panels a, b, c, and d are the left, right, even and odd arrangements, respectively. The smaller symbols are the separate signal and noise weight estimates, and the solid lines represent the optimal weights.
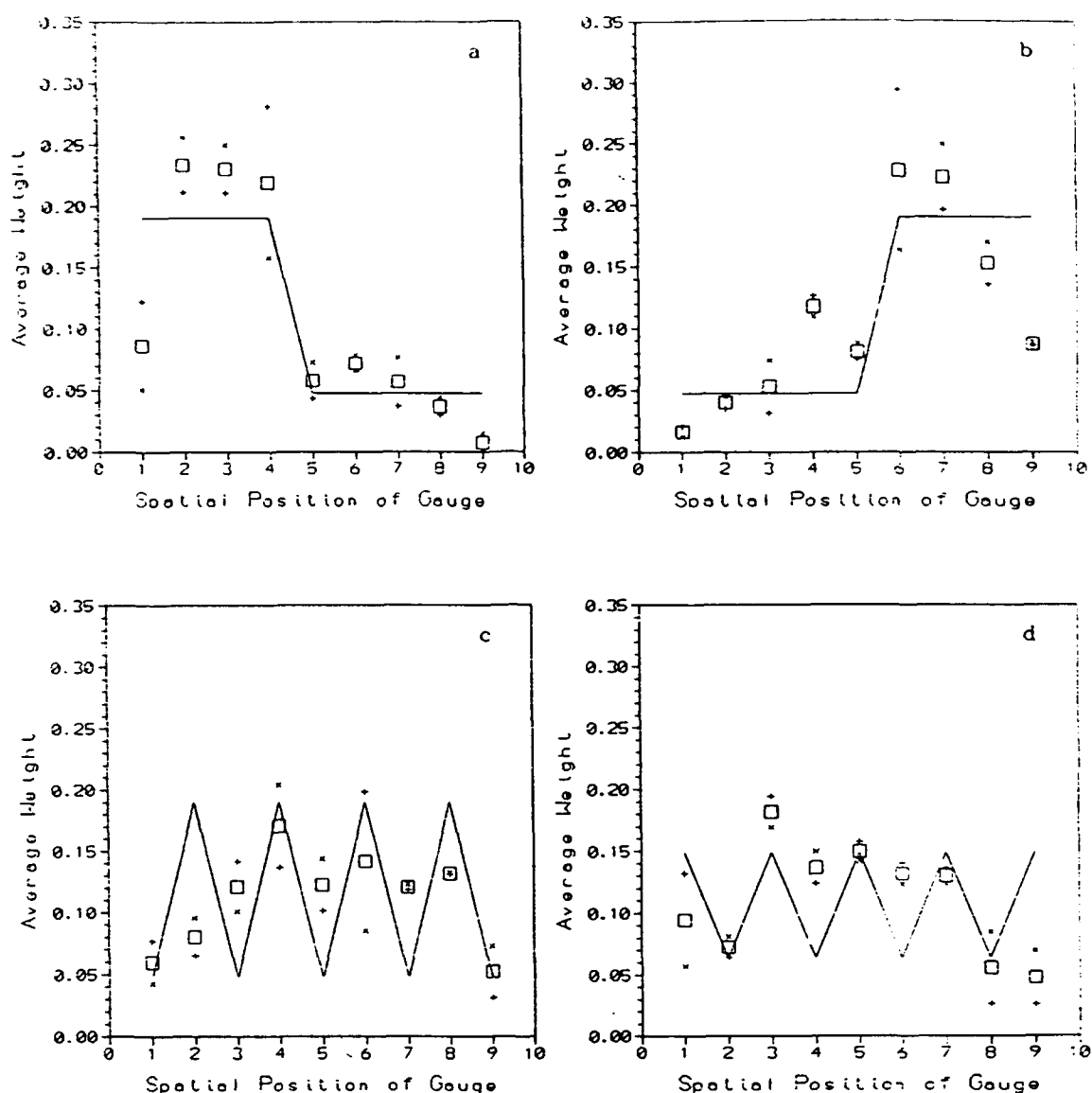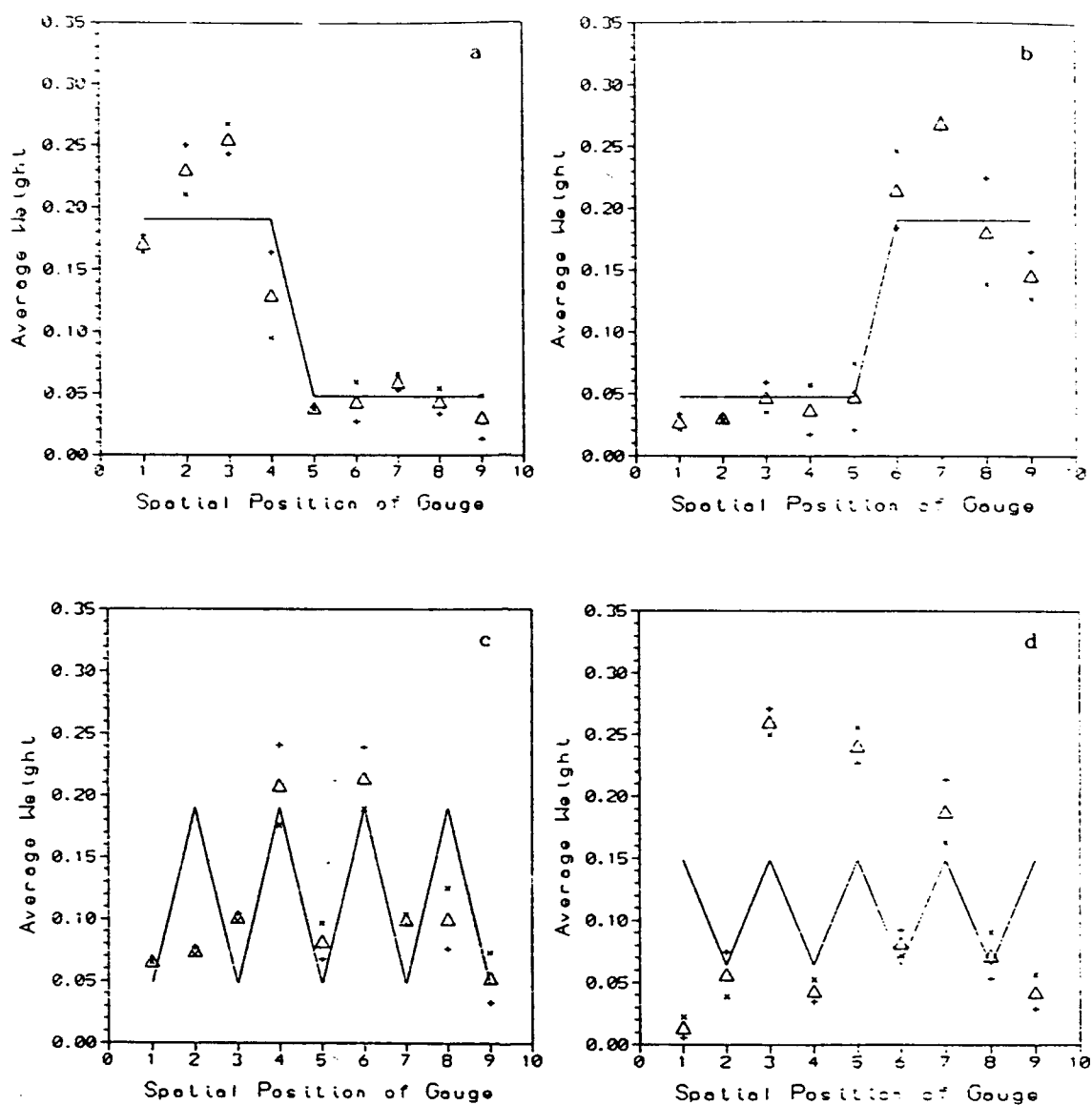
Figure 6. Subject S1's average weights for the four source
reliability arrangements in the UNcP block-type condi-
tion. Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
the solid lines represent the optimal weights.

Figure 7. Subject S1's average weights for the four source reliability arrangements in the UCM block-type condition. Panels a, b, c, and d are the left, right, even and odd arrangements, respectively. The smaller symbols are the separate signal and noise weight estimates, and the solid lines represent the optimal weights.
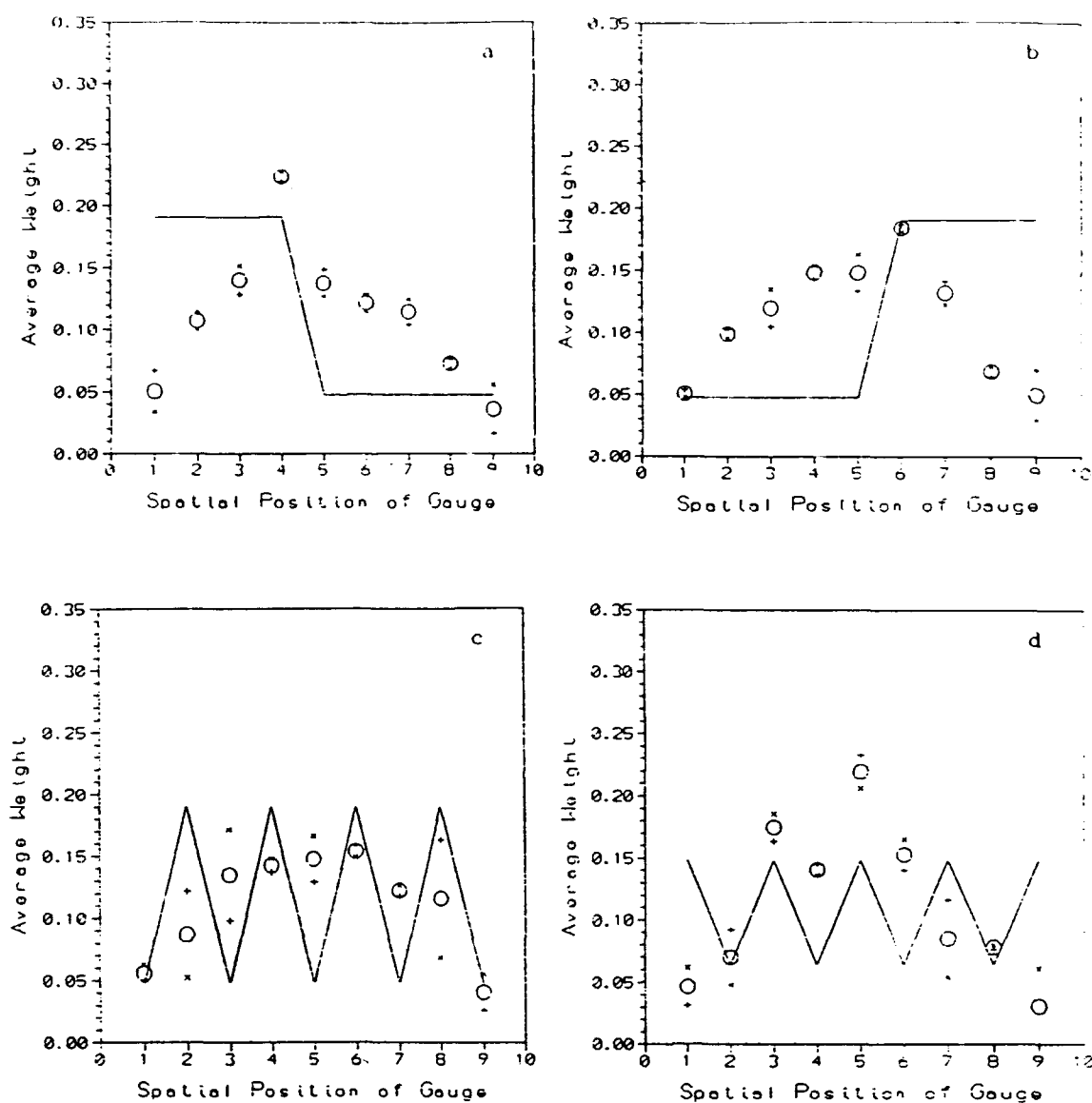
Figure 8. Subject S2's average weights for the four source
    reliability arrangements in the UNcM block-type condi-
    tion.   Panels a, b, c, and d are the left, right, even
    and odd arrangements, respectively. The smaller symbols
    are the separate signal and noise weight estimates, and
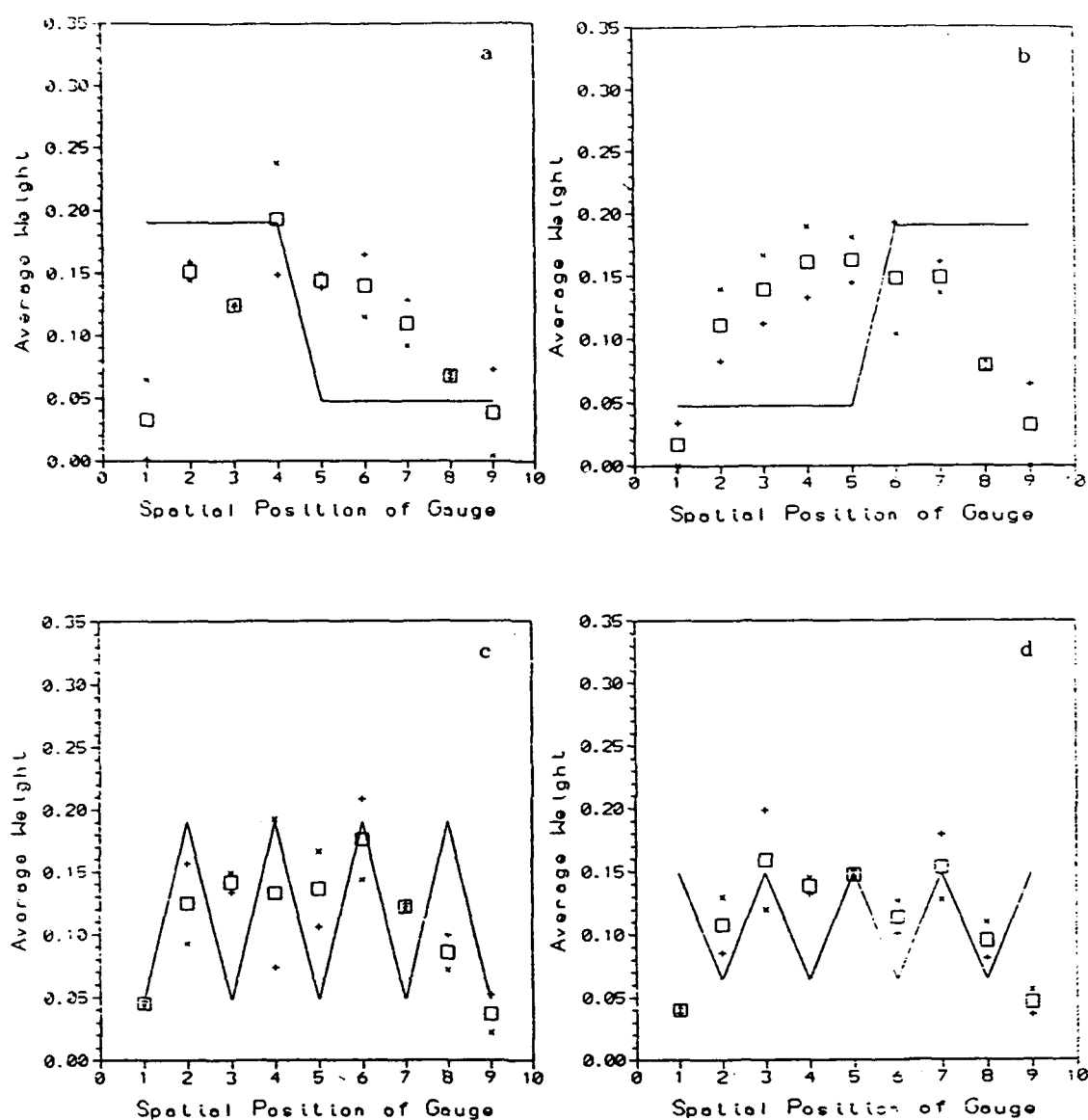   the solid lines represent the optimal weights.

Figure 9. Subject S2's average weights for the four source
reliability arrangements in the UNcP block-type condi
tion.   Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
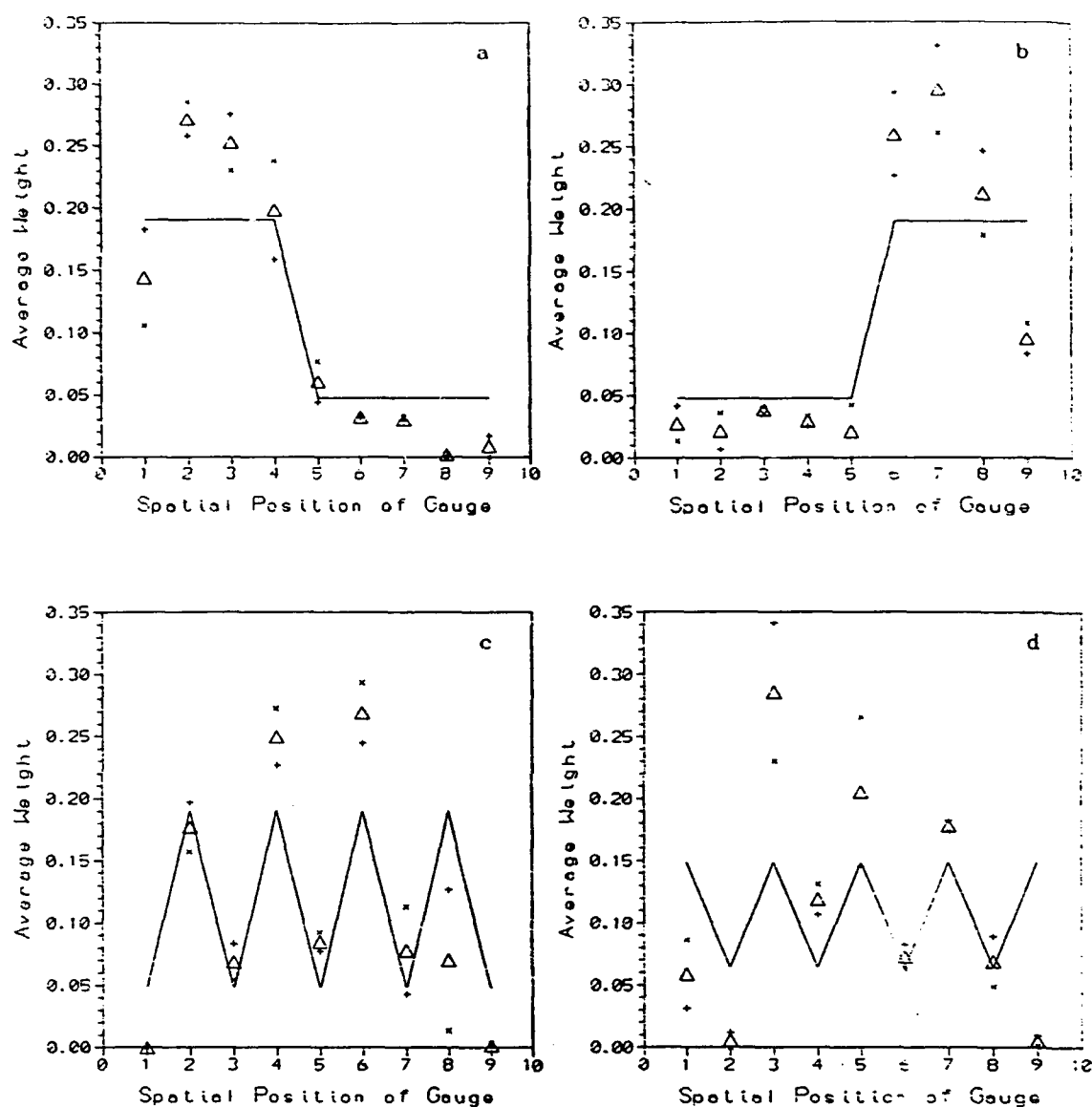the solid lines represent the optimal weights.

49

Figure 10. Subject S2's average weights for the four source
reliability arrangements in the UCM block-type condi-
tion. Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
the solid lines represent the optimal weights.

50

Figure 11.   Subject S3's average weights for the four source
       reliability arrangements in the UNcM block-type condi-
       tion.   Panels a, b, c, and d are the left, right, even
       and odd arrangements, respectively.   The smaller symbols
       are the separate signal and noise weight estimates, and
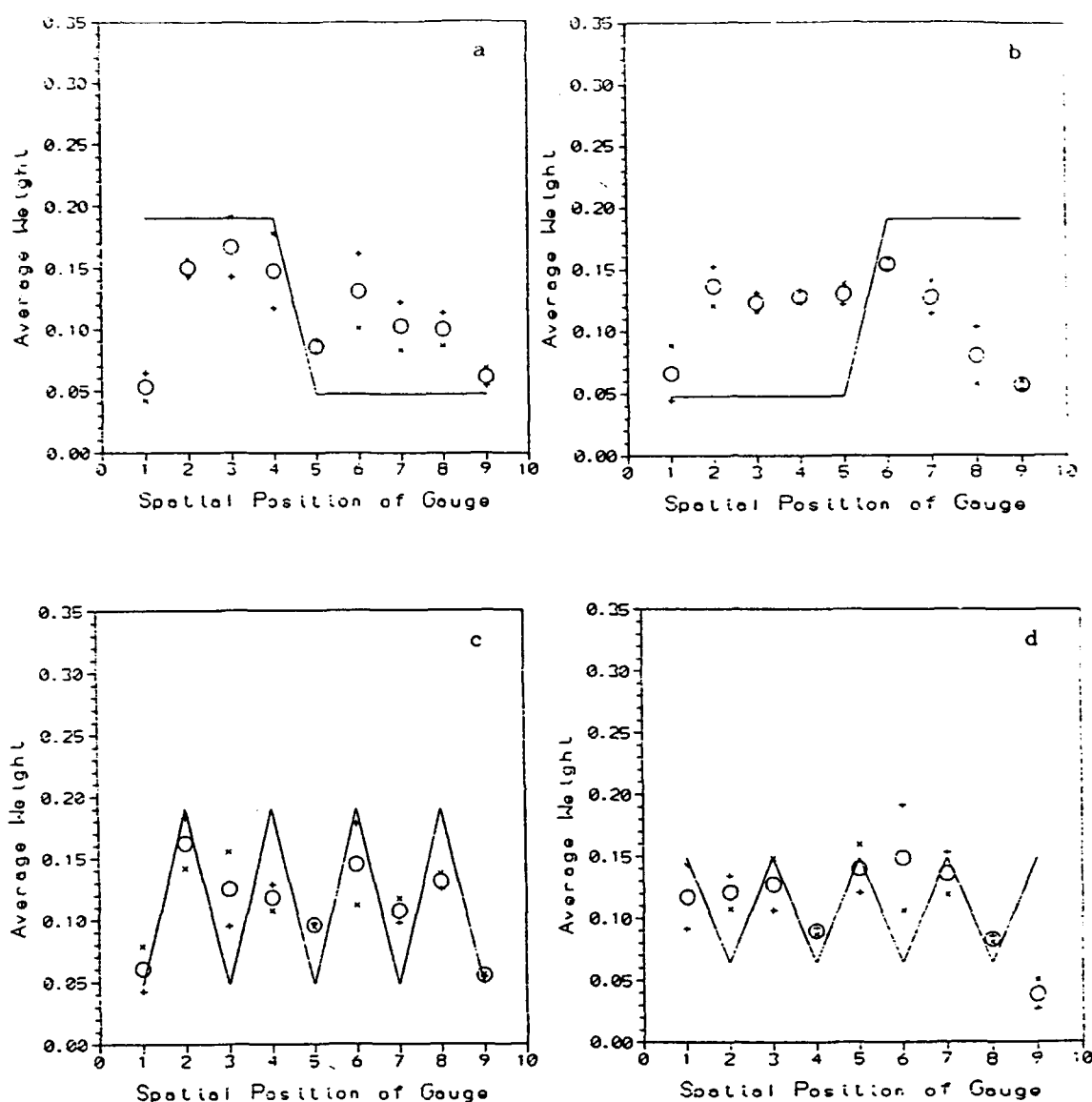   the solid lines represent the optimal weights.

Figure 12. Subject S3's average weights for the four source
reliability arrangements in the UNcP block-type condi
tion. Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
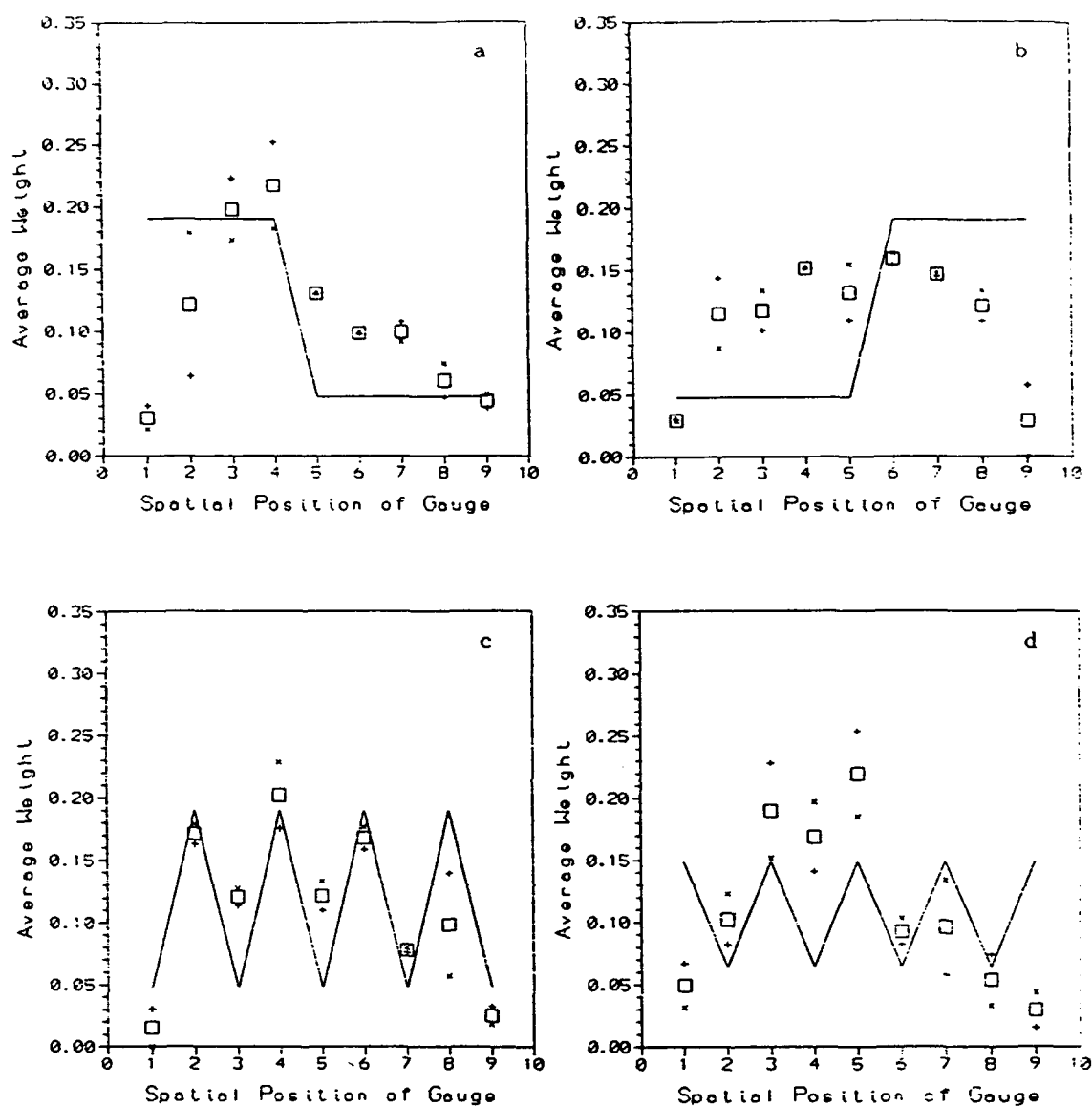the solid lines represent the optimal weights.

Figure 13. Subject S3's average weights for the four source
reliability arrangements in the UCM block-type condi-
tion. Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
the solid lines represent the optimal weights.

Figure 14. Subject S4's average weights for the four source reliability arrangements in the UNcM block-type condition. Panels a, b, c, and d are the left, right, even and odd arrangements, respectively. The smaller symbols are the separate signal and noise weight estimates, and the solid lines represent the optimal weights.
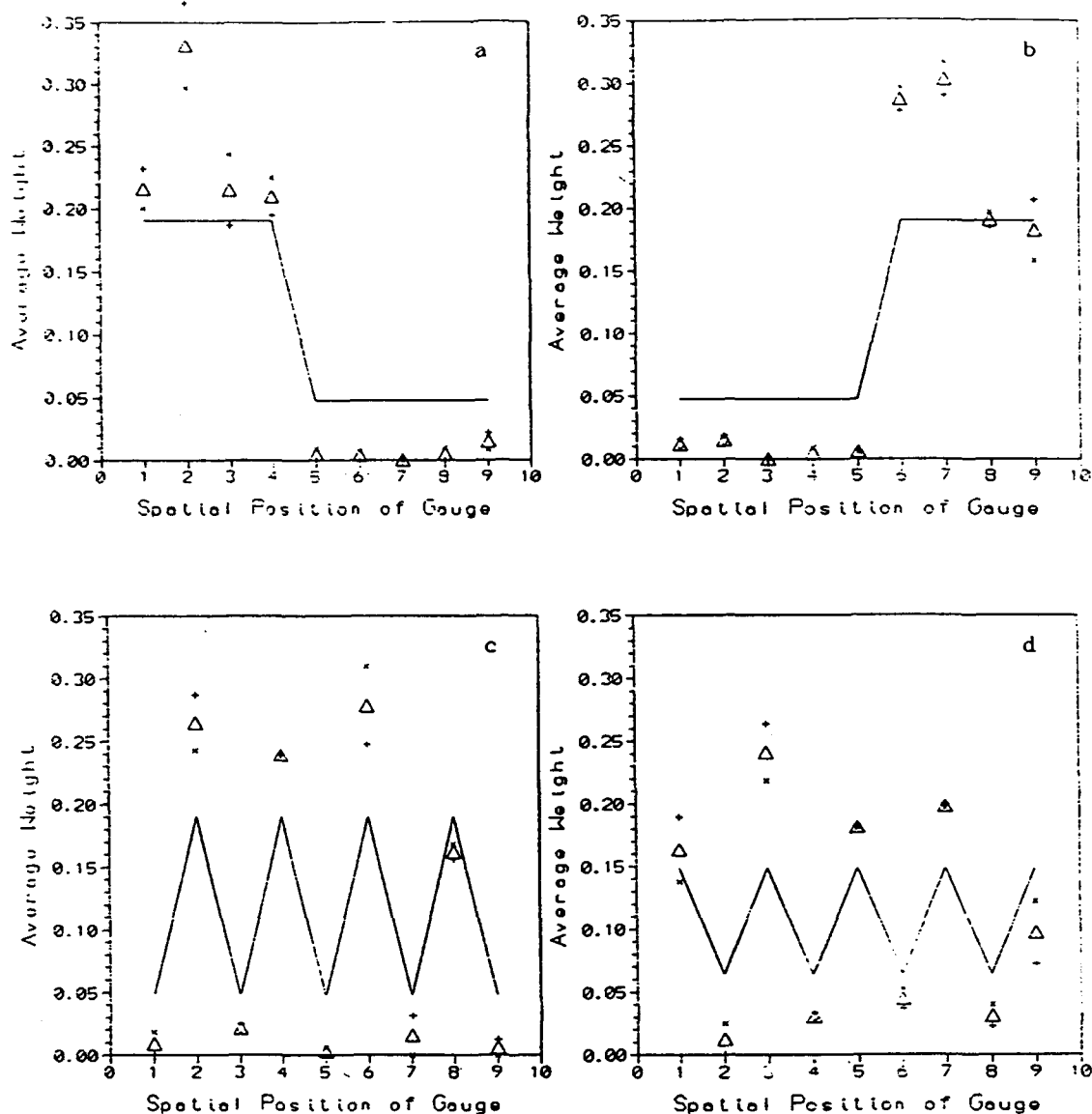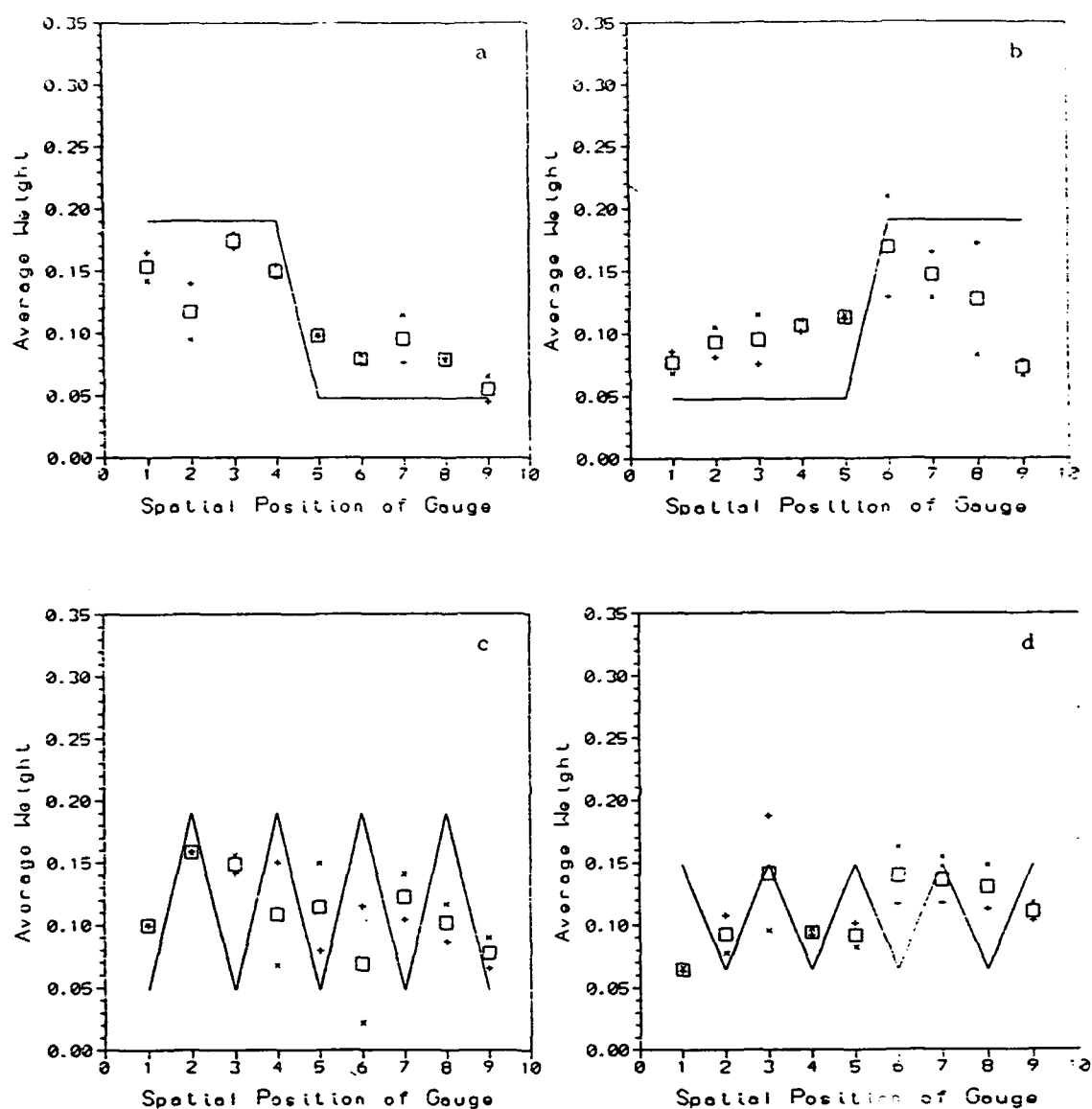
Figure 15. Subject S4's average weights for the four source
reliability arrangements in the UNcP block-type condi
tion.   Panels a, b, c, and d are the left, right, even
and odd arrangements, respectively. The smaller symbols
are the separate signal and noise weight estimates, and
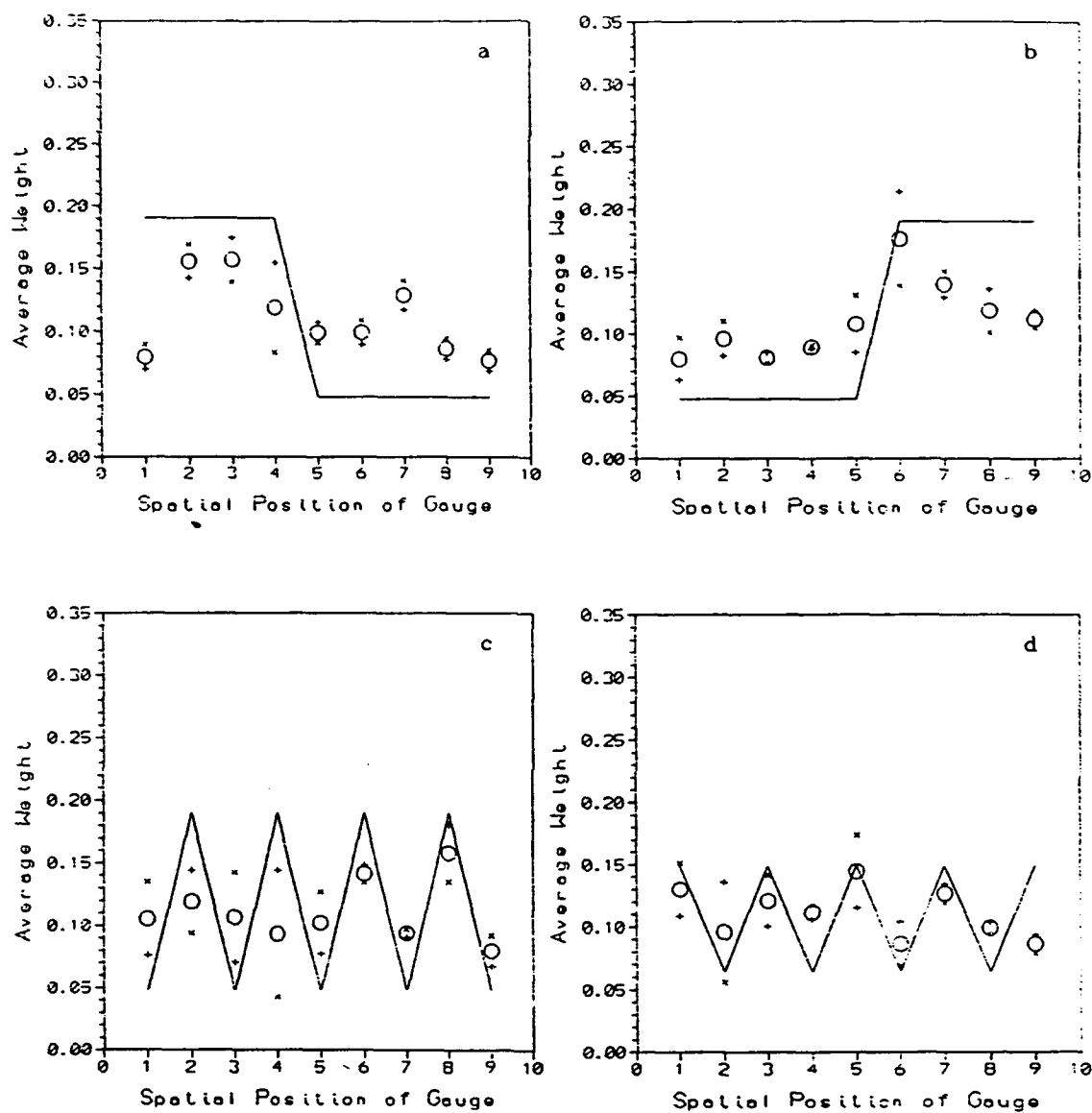the solid lines represent the optimal weights.

Figure 16. Subject S4's average weights for the four source reliability arrangements in the UCM block-type condition.   Panels a, b, c, and d are the left, right, even and odd arrangements, respectively. The smaller symbols are the separate signal and noise weight estimates, and the solid lines represent the optimal weights.
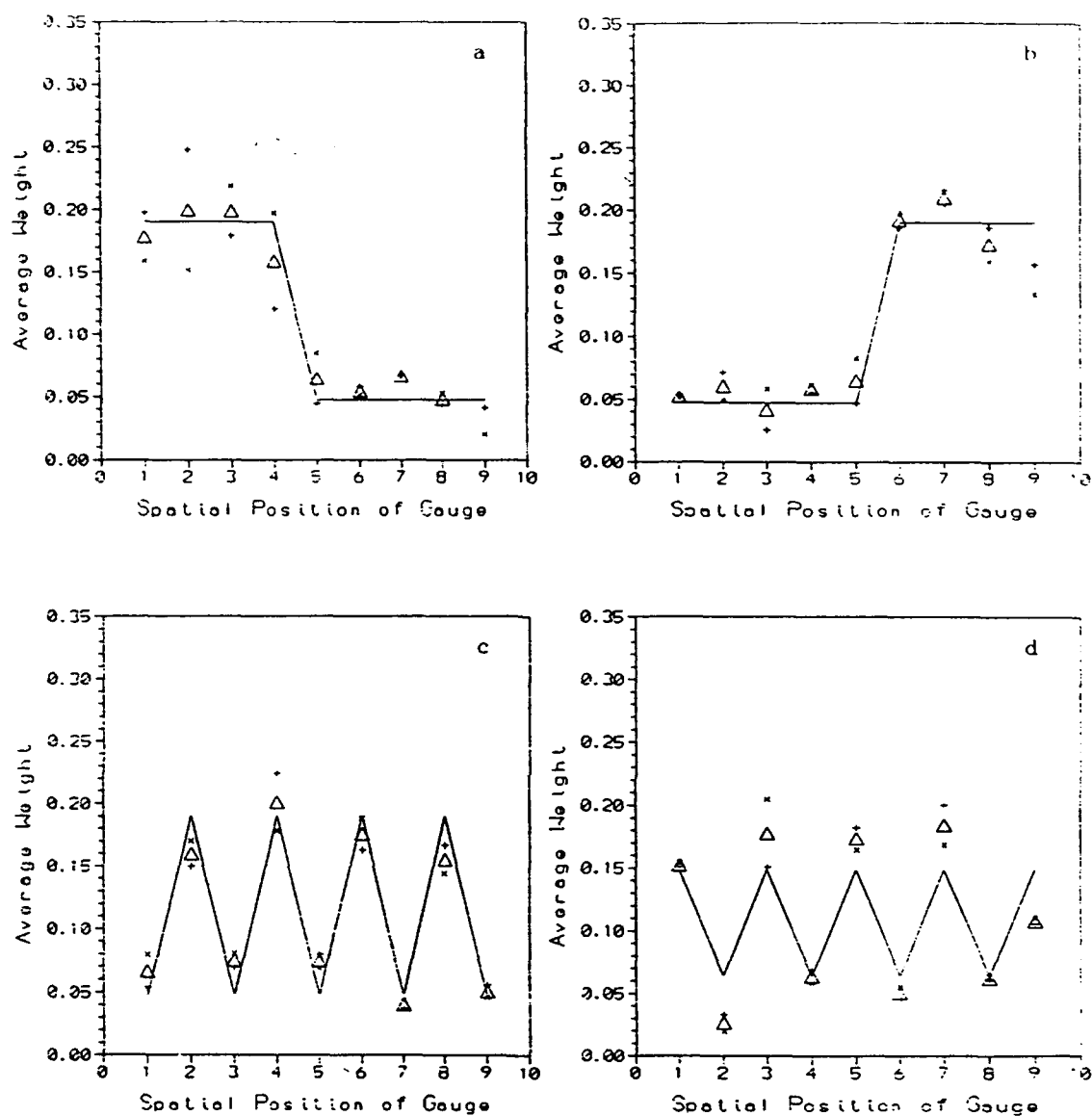
Table 6.
Weighting Efficiency Estimates for Arrangement and
the Unequal Reliability Trial-Block Conditions.

Subject	Arrangement	UNcM	UNcP	UCM
S1:	Left	0.740	0.880	0.940
	Right	0.580	0.820	0.940
	Even	0.610	0.690	0.790
	Odd	0.800	0.800	0.730
S2:	Left	0.650	0.620	0.860
	Right	0.560	0.490	0.850
	Even	0.630	0.640	0.780
	Odd	0.690	0.760	0.690
S3:	Left	0.700	0.700	0.770
	Right	0.570	0.570	0.780
	Even	0.760	0.750	0.810
	Odd	0.790	0.700	0.850
S4:	Left	0.710	0.840	0.970
	Right	0.780	0.730	0.970
	Even	0.730	0.600	0.950
	Odd	0.900	0.790	0.950
Avg:	Left	0.700	0.760	0.885
	Right	0.623	0.653	0.885
	Even	0.683	0.670	0.833
	Odd	0.795	0.763	0.805

Equal reliabilty No cue Pure block design (ENcP)
Unequal reliability No cue Mixed block design (UNcM)
Unequal reliability No cue Pure block design (UNcP)
Unequal reliability Cue Mixed block design (UCM)

Table 7.
Average weighting efficiency estimates for the
experimental variables (block-type condition,
arrangement and stimulus duration).

Stimulus Duration	Arrangement	UNcM	UNcP	UCM
150	Left	0.640	0.633	0.823
	Right	0.520	0.545	0.757
	Even	0.643	0.620	0.743
	Odd	0.673	0.610	0.635
400	Left	0.668	0.730	0.845
	Right	0.585	0.675	0.890
	Even	0.660	0.658	0.815
	Odd	0.728	0.730	0.740
800	Left	0.700	0.760	0.885
	Right	0.623	0.653	0.885
	Even	0.683	0.670	0.833
	Odd	0.795	0.763	0.805
AVG	Left	0.669	0.708	0.851
	Right	0.576	0.624	0.844
	Even	0.662	0.649	0.797
	Odd	0.732	0.701	0.727

Equal reliabilty No cue Pure block design (ENcP)
Unequal reliability No cue Mixed block design (UNcM)
Unequal reliability No cue Pure block design (UNcP)
Unequal reliability Cue Mixed block design (UCM)

shows higher weighting efficiency than the other two conditions, UNcM and UNcP, for each arrangement and stimulus duration, except when the most reliable elements were found in the Odd positions. Table 6 indicates that all four observer show the same pattern of effects at the 800 ms stimulus duration. They tended to show a little more variability, but similar patterns across subjects at the two shorter stimulus durations.

As with the d'_{obs} data we see a slight, but not significant, advantage for grouped arrangements in the UCM condition, and for the odd arrangement in the UNcM condition relative to the other arrangements. The interaction between stimulus duration and arrangement found in the d'_{obs} data was not supported by differences in observers weighting efficiencies.

From table 6 we also see that the less practiced observers (S1, S2, and S3) tend to show the lowest weighting efficiency in the two non-cued conditions, UNcM and UNcP, for the right arrangement. The same is true for the shorter durations. However, once the cue was introduced, efficiency was indistinguishable for the weighting efficiency for the left arrangement which was consistently fairly high. Finally, as far as the observers' residual efficiency, η_{obs}, were concerned, the only consistent difference among the observers was a tendency for residual efficiency to be higher for the equal reliability condition relative to the three unequal reliability block-type conditions. The stimu-

lus duration by arrangement interaction found in the observers' d'$_{obs}$ measures was not driven exclusively by changes in η_{wgt} or η_{noise}, but rather by a combination of these effects.

Discussion

The primary goal of this investigation was to determine the effects of selected display factors in directing observers' attention to informative sources. There was an overall improvement in observer performance as stimulus duration increased; this was mainly a function of improved weighting efficiency. In general, when no additional cues to source reliability were available, weighting efficiency was greatest when sources were equal rather than unequal in reliability. There was a tendency for better performance when more reliable sources were grouped, rather than distributed in the cued block-type condition, especially when the stimulus duration was at least 400 ms. Alternatively, in the no cue block-type conditions performance was highest when observers were presented a distributed odd arrangement. These differences in performance were mostly due to differences found in the observers' weighting efficiency measures.

When sources have to be prioritized in terms of the underlying statistical properties of the information, as in this study, people are limited by both their ability to estimate these properties (e.g., variability of the source) and by their ability to then weight the sources accordingly. The observers' relatively poor performance in the non-cued

conditions may have been due to their inability to estimate the variability of the sources when this information was relevant to their decisions (e.g., the UNcP condition). That is, the observers may not have been sensitive to the trial-to-trial variability of the sources. The improvement in weighting efficiency, given the luminance cue in the UCM condition, indicates greater attention or weight directed toward these elements. It does not necessarily suggest that observers have improved sensitivity to the differences in element reliability. To test this possibility, observer S4, contributed data to an additional condition in which gauge luminance was inversely related to gauge reliability.

Figure 17 shows subject S4's average weight estimates for the two luminance cue conditions, UCM and reverse cue, and the four reliability patterns, Grouped-Left and -Right and Distributed-Even and -Odd. The circles represent a direct relationship between gauge luminance and reliability, the UCM condition. The triangles represent the reverse cue condition. The weights estimated for these two conditions are nearly identical. This was reflected in the weighting efficiencies. The largest difference between the two luminance cue conditions in weighting efficiency was 0.02; for the Grouped-Left pattern $\eta_{wgt} = 0.97$ for the UCM condition, and $\eta_{wgt} = 0.95$ for the reverse cue condition. Thus, at least on the initial trials in a given block the observer had to be sensitive to differences in element variability to detect which elements were more reliable.
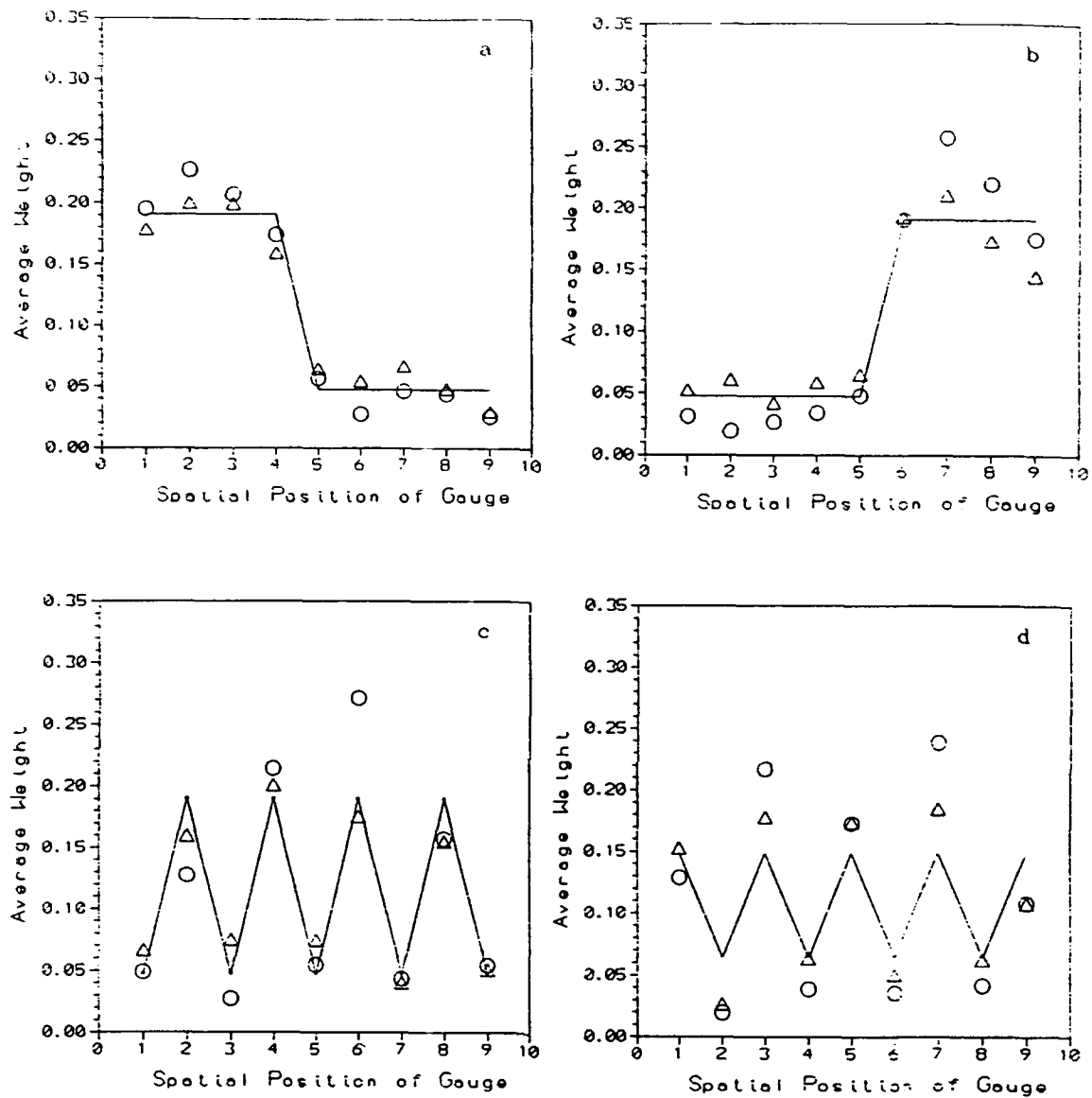
Figure 17. The average weights for subject S4 in the UCM and reversed cue conditions.

However, there remains a question as to whether or not observers were actually using trial-to-trial variability of the sources to make their decisions. Some of the observers showed a fair amount of accuracy in weighting sources according to reliability in the UNcM condition where trial-to-trial variability could not be used to identify which sources were more reliable. One possible explanation for this performance is that observers were sensitive to patterns produced by gauge markers when the variability was low.

On a given trial, the markers of the high reliability gauges tended to fall at a common vertical position in the display, causing them to line up as shown in figure 18. Thus, this pattern may have drawn observers attention, helping them to identify which sources were more reliable. After questioning the observers about strategies used on these trials, it was confirmed that observers were sensitive to such patterns in the display. Based on this evidence alone it is not conclusive that observers were sensitive to the underlying variability of the sources. Rather it is more probable that cues provided and the patterns inherent in displays helped observers to weight sources according to reliability.

Finally, the especially good performance for the odd arrangement raises another question. Why is performance so good for the odd arrangement in the non-cued conditions? This advantage may be due to the unique characteristics of
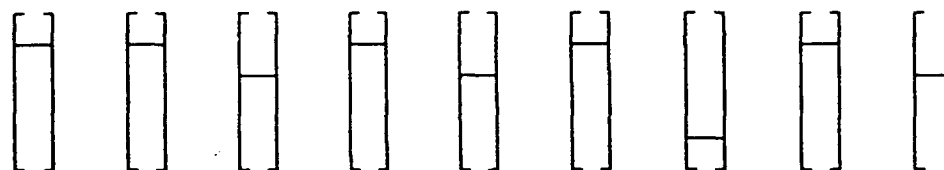
Grouped-Left High
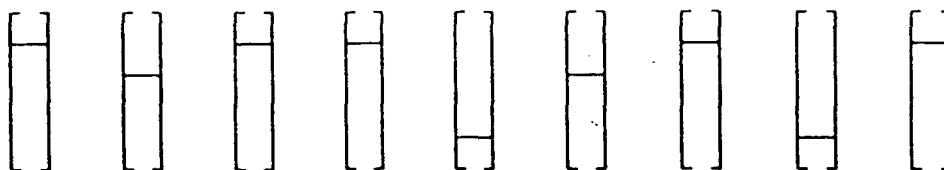
Grouped-Right High
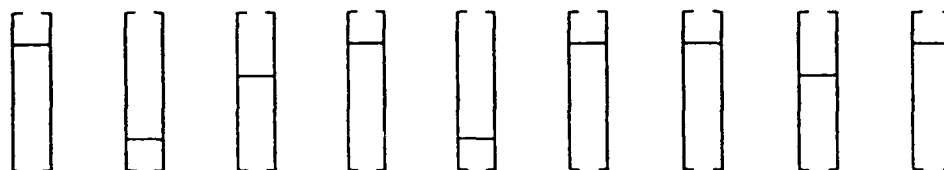
Distributed-Even High
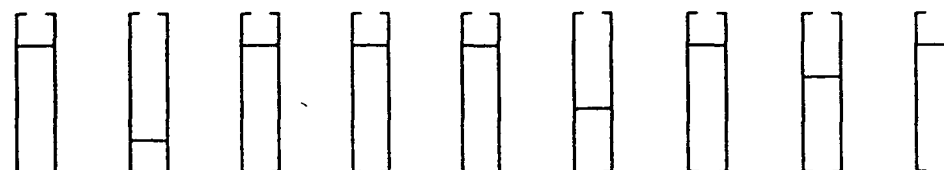

Distributed-Odd High

Figure 18. Demonstration of the possible patterns for
    four arrangements which observers may have used to
    identify the more reliable sources.

the arrangement. The odd arrangement was the only condition with five sources, rather than four, which were high in reliability. As a result, it had more high reliability sources distributed throughout the visual field, and it was the only arrangement that had a source high in reliability located at the fixation point. Secondly, in order to maintain equal levels of predicted ideal performance, there were smaller differences between the variances for high and low reliability sources. Thus, based on both of these factors, this condition most closely approximated an equal reliability condition. If observers resorted to weighting sources equally when they are under time stress or unable to identify which sources are reliable, this strategy would prove most useful in the odd condition.

In conclusion, from this study we can state that when observers have to utilize information from multiple subordinate displays they are relatively inefficient at identifying differences in reliability among the displays. However, there is improved efficiency when the display elements are coded by luminance. The assumption is that this cue, and possibly other cues, help observer's to prioritize the information by indicating where attention should be directed. Additional assistance may be gained by organizing the displays such that sources similar in reliability are proximate to one another.

EXPERIMENT 2

Introduction

This study continues the investigation of observers'
ability to use independent visual information sources in
forming detection decisions. Since humans are often
required to make decisions under time stress in many real
world settings, researchers have been interested in identi-
fying means of coding visual information to reduce potential
errors and optimize human performance. One approach is to
assist performance by creating display codes which capital-
ize on our knowledge of human sensory and perceptual mecha-
nisms. For instance, Woods, Wise, and Hanes (1981) reduced
the complexity of integrating multiple independent sensor
values to form a detection decision by combining display
elements into a single geometric form. This allowed human
monitors to use shape distortions to identify important
system states.

The primary concern of this study was to determine
whether two factors related to display element arrangement
affect observers' detection decisions. The first factor
concerns the influence of emergent features on observers'
detection decisions. Emergent features are defined as
properties that arise from the configuration of "simple"
elements that are not identifiable in any given element

(Treisman, 1986). For instance, if the elements are repre-
sented by three line segments, then depending on the
arrangement chosen we could create one of the forms shown in
figure 19. Particular element arrangements produce features
such as angles and intersections which are not observable
given the individual lines. Moreover, some element arrange-
ments produce global features which are recognizable
objects. For instance, the first and last element arrange-
ments in Figure 19 do not possess as strong an object quali-
ty as the triangle found in the middle.

There is mixed evidence in the object perception litera-
ture regarding whether emergent, object-like, features of
simple element arrangements facilitate or hinder detection
of the underlying elements. Some studies have found evi-
dence suggesting an "object-superiority effect." That is,
when a target feature (e.g. a line segment of a given orien-
tation) is embedded in a contextual pattern, observer detec-
tion performance is facilitated when the target feature and
context form a recognizable object (Weisstein & Harris,
1974). Similarly, Ankrum and Palmer (1991) found that
observers were better at detecting differences between two
element arrangements which formed objects than between
element arrangements in which one was an object and the
other was part of an object. This enhanced detectability
may be related to the familiarity of the organized pattern
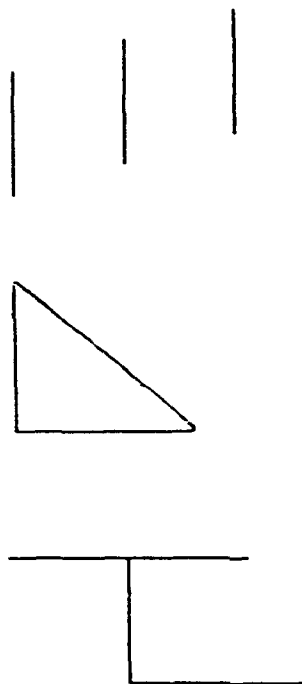(Purcell & Stewart; 1991), similar to the "word superiority

Figure 19.   Example of three line-graph arrangements
demonstrating an emergent feature.   The middle figure
produces an emergent object-like feature.

effect" where it is easier to detect a specific letter in a word than in a nonword (Reicher, 1969).

Others have found contradicting evidence. Pomerantz (1981) points to evidence suggesting that when elements perceptually group, the emergent feature created by the configuration may be more perceptually salient and selective attention to the underlying elements may be impeded (Pomerantz & Schwaitzberg, 1975). Similarly, Navon (1977) found Stroop interference of global configurations on subjects' processing of local elements, but not the opposite. Bennett and Flach (1992) summarize the results from a number of studies which have applied this concept to real world settings. These studies indicated that an emergent object-like property had no affect on, or adversely affected, performance in tasks requiring focused attention to the individual elements. However, they pointed out that performance had been facilitated by the same emergent features when the detection task required information integration.

One hypothesis, tested in earlier studies, stated that the magnitude of the performance advantage in integration tasks, relative to selective attention tasks, depended upon the degree to which the element configuration produced an object-like feature (Carswell & Wickens, 1987; Wickens & Andre, 1988). The degree of "objectness" depended on whether or not the element configuration possessed an enclosed contour (Wickens & Andre, 1990). However, later evidence suggested that the important factor was whether or not an

emergent feature carried important information about the underlying state, rather than simply the object quality of the configuration. For instance, Buttigieg and Sanderson (1991) found that object displays did not always produce the best performance in integration tasks, whereas "well-mapped" emergent features did.

However, recent investigations, designed specifically to address the relationship between the configuration and the task, have produced mixed results. Some researchers have found support, suggesting that performance was better when there was a strong relationship between some property of the emergent feature and the decision statistic than when this relationship was weak (Bennett, Toms, & Woods, 1993; Mitchell & Biers, 1992; Schmidt & Elvers, 1992). Others researchers did not find a performance advantage for emergent features that were "well-mapped" to the decision task (Sanderson, Haskell, & Flach, 1992).

The second factor addressed in this study concerns the importance of the relationship between the emergent feature and the optimal decision statistic for the task. One nice characteristic of the Theory of Signal Detectability (TSD, Green and Swets, 1966) paradigm is that we can mathematically specify the optimal decision statistic in a detection task. The optimal decision statistic is the likelihood ratio or some value which is monotonically related to the likelihood ratio. If there is a direct relationship between an emergent property in the display (e.g., size or area) and

the optimal decision statistic, it is expected that the observers can use changes in the magnitude of this emergent property to make their decisions.

In this study, observers were presented four independent informational sources coded as graphical elements in a visual display. Six display codes were constructed from a combination of two factors, 1) whether or not the display element arrangement produced an emergent feature, and 2) whether or not the emergent feature had some property that was directly related to the optimal decision statistic (for the Yes/No task). Observers used this information to perform either a Yes/No task or a 4AFC task.

The magnitude of a given source was determined by a normal random variable which depended on the underlying state (Signal or Noise). For signal events, the source values were selected from a distribution with a mean of μ_s and a standard deviation of σ_s. The values of noise sources were drawn from a distribution with a mean of μ_n and a standard deviation of σ_n, where $\mu_n < \mu_s$ and $\sigma_s = \sigma_n = \sigma_e$. In the Yes/No task, observers had to decide whether all source values presented on a given trial represented either a signal state or noise. In the 4AFC task, the observer had to decide which of four elements represented the signal event.

If emergent features facilitate the processing of the underlying elements, then we may find more efficient decision making performance for both Yes/No and 4AFC decision

tasks when such features are present. Alternatively, if these emergent features are processed faster than the underlying elements, then decision tasks which require sensitivity to the underlying elements may be hindered by display codes that possess these features. For instance, in an equal reliability Yes/No task, sensitivity to the separate underlying elements is of less importance to performance than it is in a 4AFC task. As a result, if the emergent feature hinders observer sensitivity to the underlyii.; elements, performance on a 4AFC task may be less efficient when the information is arranged to form an emergent feature than when the element arrangement does not posses this feature.

With respect to the second factor, if the relationship between the magnitude of some emergent feature property and the optimal decision statistic is important then we should see a performance advantage when this relationship is present. In the current study, this relationship was coded only for the Yes/No decision task. Two of the six display arrangements produced an emergent feature in which the width or the area of this feature was directly related to the optimal decision statistic for the Yes/No task. Thus, it was expected that detection performance would be facilitated in the Yes/No task given these two display codes relative to the other codes which do not possess this relationship.

Finally, the object-like quality of the emergent feature was also tested in this study. Some of the element

arrangements produced emergent features which possessed an enclosed contour, whereas others did not. If the configural property of the display arrangement is an important factor (Carswell & Wickens, 1987), then display arrangements which possess this property may be more likely to show the expected emergent feature effects than arrangements without an enclosed contour.

Method

Subjects

Three of the four subjects who participated in the first study also contributed data in this study. All observers were paid an hourly wage plus a bonus based on performance.

Apparatus and Stimuli

Observers were seated in a sound isolated booth approximately 27 inches away from a 10.5 inch color monitor (EGA). The monitor was set for maximum contrast. Intensity was set at approximately 100 cd/m² measured from a uniform white field covering the monitor. On a given trial, one of six arrangements of four graphical elements was presented on the monitor against a gray grid. The maximum horizontal and vertical visual angles were 13.5° and 4.5°, respectively. (The separate measures of visual angle for each of the displays are found on figures 20 and 21.) The values depicted by the graphical elements were either drawn from a _signal_ or _noise_ distribution, depending on the trial and task. The parameters of the _signal_ and _noise_ distributions

were $\mu_s = 50$ and $\sigma_s = 10$ and $\mu_n = 40$ and $\sigma_n = 10$, respectively. Element magnitude ranged from 1 to 99.

Figures 20a - 20c depict three line-graph display codes in which element magnitude was coded by the length (number of pixels) of a horizontal line segment. Figure 20a represents the linear likelihood (LIN-LR) arrangement. In the LIN-LR arrangement, there was a fixed separation between the end of one line segment and the beginning of the next. Thus, in the Yes/No task the total length of the display produced by the separate segments varied directly with the likelihood ratio. Figures 20b and 20c represent two versions of the linear non-likelihood arrangement. In both cases, the onset of each line segment began at a specified location in the visual field; total display length did not vary. In one case the elements were arranged horizontally (LIN-NL), and in the other case they were arranged in a square (LSQ-NL), to control for differences in visual angle.

Figures 21a - 21c depict three angular displays. The angle formed by two line segments in a given quadrant was directly related to the magnitude of the underlying element. One end of each line segment was fixed in position on the display. The opposite ends of the two segments joined to form an angle. The small arrows in figures 21a - 21c designate the angles being described. The size of the angle varied with element magnitude as follows:

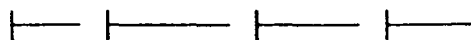$$\text{Angle} = 270 - 2\tan^{-1}\left((100-x_i)/x_i\right), \qquad (22)$$

where x_i is the magnitude of the ith element. In figure

a.) Linear Likelihood (LIN-LR) Arrangement

Yes/No Signal Trial

├──── ├────── ├──── ├────
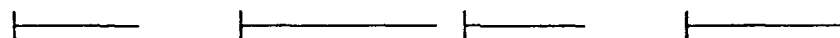
Yes/No Noise Trial

├── ├───── ├──── ├───
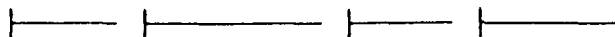
Average Visual Angle: Horizontal = 6.75°
                      Vertical   = 1.6°

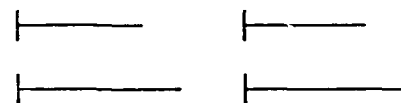b.) Linear Non-Likelihood Horizontal (LIN-NL) Arrangement

Yes/No Signal Trial

├──── ├────── ├──── ├──────

Yes/No Noise Trial

├── ├─── ├──── ├──

Average Visual Angle: Horizontal = 11.25°
                      Vertical   = 1.6°

c.) Linear Non-Likelihood Square (LSQ-NL) Arrangement
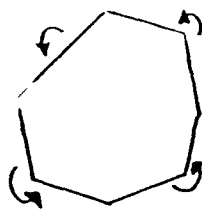
Yes/No Signal Trial   Yes/No Noise Trial

├──── ├───   ├── ├────

├───── ├────   ├── ├─

Average Visual Angle: Horizontal = 6.2°
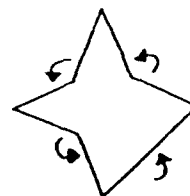   Vertical   = 6.3°

Figure 20. Line graph display codes. Figures a, b and c
    are the Linear Likelihood (LIN-LR), Linear Non-Likeli-
    hood (LIN-NL), and Linear Non-Likelihood Square (LSQ-NL)
    displays. Each display was presented in front of a gray
    grid and the visual angle subtended is listed at the
    bottom of each figure.

a.) Object Likelihood (OBJ-LR) Arrangement

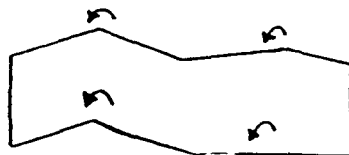Yes/No Signal trial   Yes/No Noise trial

Average Visual Angle: Horizontal = 6.3°
                    Vertical   = 6.3°

b.) Object Non-Likelihood (OBJ-NL) Arrangement

Yes/No Signal trial   Yes/No Noise trial

Average Visual Angle: Horizontal = 6.3°
                    Vertical   = 6.3°

c.) Angular Non-Likelihood (ANG-NL) Arrangement

Yes/No Signal trial   Yes/No Noise trial
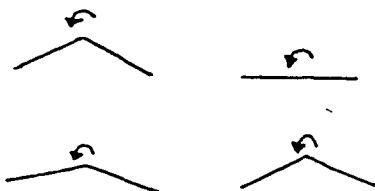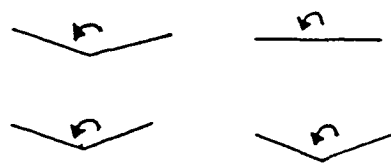
Average Visual Angle: Horizontal = 6.3°
                    Vertical   = 6.3°

Figure 21. Angular display codes. Figures a, b and c are the Object Likelihood (OBJ-LR), Object Non-Likelihood (OBJ-NL), and Angular Non-Likelihood (ANG-NL) displays. Each display was presented in front of a gray grid and the visual angle subtended is listed at the bottom of each figure.

21a, the smaller element values correspond with angles which point toward the center of the display. In figures 21b and 21c, smaller element values correspond with angles which point downward.

In two of the figures (21a and 21b) the elements were arranged to form an enclosed contour, producing object-like shapes. In figure 21a the element arrangement produced an object in which the area was directly related to the likelihood ratio for the Yes/ No task (OBJ-LR). That is,

$$Area = \sum_{1}^{4} 100x_{i}. \tag{23}$$

Figure 21b represents an object-like display that does not have a property related to the optimal decision statistic (OBJ-NL). Finally, figure 21c depicts the ANG-NL display which is identical to the OBJ-NL display; however, it does not posses a continuous enclosed contour.

Procedure

For the Yes/No task, observers were told to make their decisions based on the average magnitude of the gauges, and to rank the likelihood that the evidence represented a signal by using the "4", "3", "2" and "1" keys of a 101-key keyboard, where "4" represented "very sure it was a signal" and "1" represented "noise". Again, "1" and "2" responses were identified as noise, and "3" and "4" responses were identified as signals in the analyses. On 4AFC task, observers were told to identify which of the four gauges had the greatest magnitude, and thus, represented a signal event.

Subjects indicated the signal location by using either the same keys found in the Yes/No task, or the "Ins", "Home", "Del" and "End" keys, depending on the element arrangement.

The trial sequence was the same as the first study (see Figure 4). However, instead of nine line-graph gauges, one of the six display arrangements described above were presented for a stimulus duration of 200 ms. (The stimulus duration was controlled in the same manner as described in experiment 1.) Each observer received eight blocks of practice for each of the display arrangements and tasks before the experimental data was collected. For both the practice and experimental trials, the display arrangements were randomly presented and each observer received a different random order. In a given session, a subject ran through eight blocks of the Yes/No task and eight blocks of the 4AFC task, and they contributed data to eight blocks of one task before beginning the next task. The order of the tasks alternated across experimental sessions.

Since performance in the initial experimental blocks was nearly ideal, a random noise pattern was added to the displays to degrade performance. The random noise pattern consisted of 750 white spots, two pixels in width. On each trial, the locations of the spots were randomly determined. Thus, the pattern varied across trials. This noise pattern overlaid the graphical elements and background grid, such that the extent of the random noise was confined to the vertical and horizontal dimensions of the background grid.

Approximately 12% of the grid region was covered by the random noise pattern.

Subjects performance in the Yes/No task was poorer with the random noise pattern than without. Overall efficiency dropped approximately 16% when the random noise was added. The size of the differences increased. However, the general pattern of effects did not change. Thus, the experimental trials included the random noise pattern.

Results

Yes-No Task

The observers' accuracy (d') and mean reaction time measures for the six display arrangements in the Yes/No task are shown in figures 22 and 23, respectively. Three panels, a-c, in each figure represent the individual subjects' data, and the fourth, d, is the three observers average data. Each of the subjects d' and reaction time measures are based on eight blocks of 100 trials, and the error bars represent one standard error of the mean.

Separate repeated measures ANOVAs were performed on the observers' d' and reaction time measures, collapsed across the eight trial blocks. There was an effect of type of display arrangement ($F(5,10)=11.263$, $p \leq 0.001$) on observer accuracy (d's). Subsequent analytic comparisons, using the pooled variance as the error term, indicated that performance was greater for element arrangements that produced emergent features ($F(1,10)=23.2$, $p \leq 0.001$) relative to those that did not have such features. Whether or not the emergent feature produced an enclosed contour did not influence

Figure 22. The observers' average performance (d') measures for the six arrangements in the Yes/No task. Panels a-c represent the individual subjects and panel d is the average data. The error bars are one standard error of the mean.

80
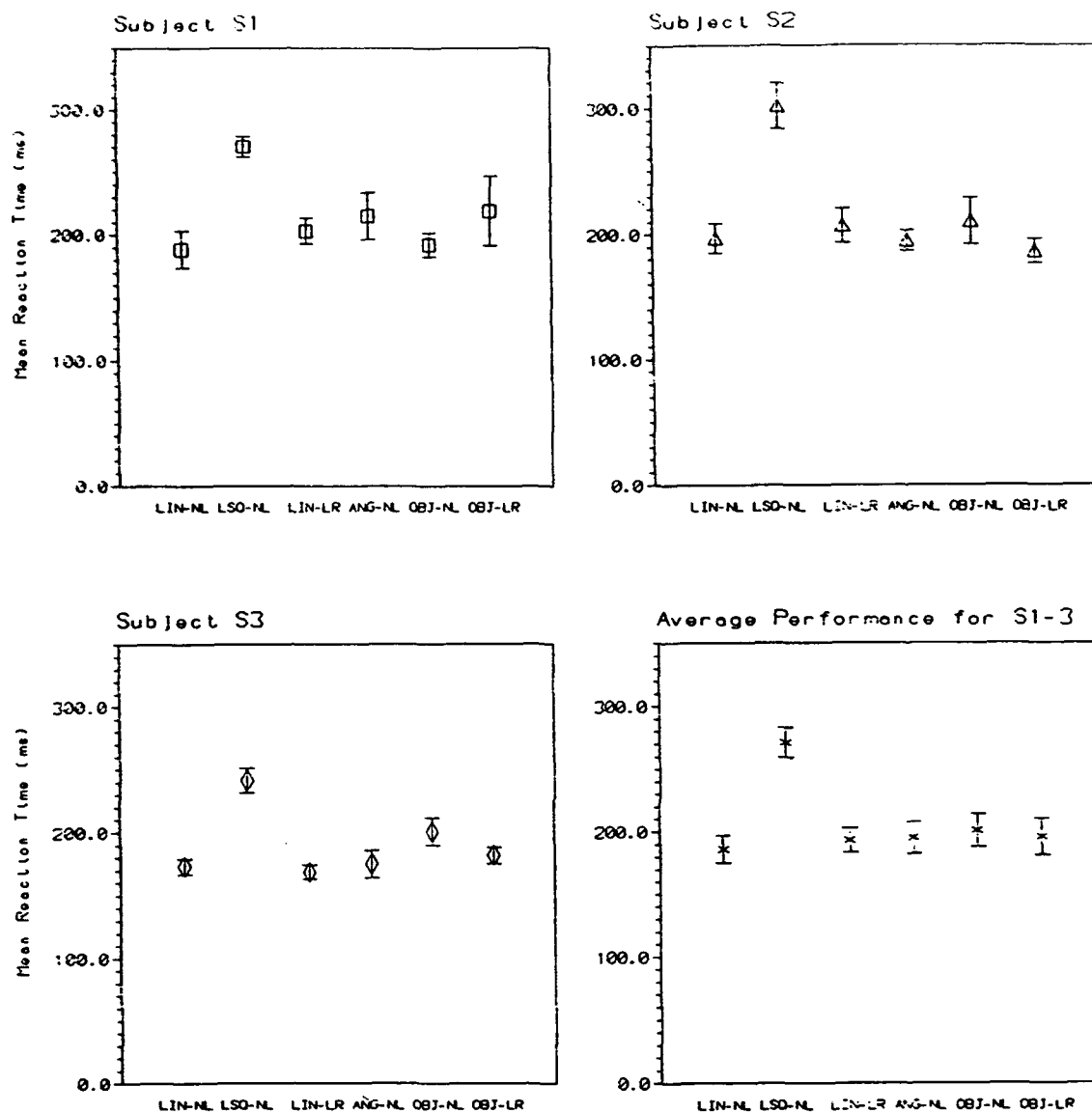
Figure 23. The observers' average reaction times measures
(measured from the offset of the mask to response) for
the six conditions in the Yes/No task. Panels a-c
represent the individual subjects and panel d is the
average data. The error bars are one standard error of
the mean.

pe.formance. However, for the non-emergent feature dis-
plays, performance was significantly better ($F(1,10)=30.8$, p
≤ 0.001) for the ANG-NL display relative to the LIN-NL and
LSQ-NL displays. This performance advantage among the non-
emergent feature displays may be a function of the underly-
ing angular element code, especially since the difference
between the ANG-NL and OBJ-NL display arrangements was not
significant.

The effect of the likelihood ratio manipulation was
significant only for the line-graph displays
($F(1,10)=21.499$, $p \leq 0.001$). Performance in the LIN-LR condi-
tion was better than performance given the other two linear
displays, LIN-NL and LSQ-NL. Finally, there was a signifi-
cant difference among the observers' reaction time measures
($F(5,10)=13.4$, $p \leq 0.001$) for the separate element arrange-
ments. Reaction time was slower given a LSQ-NL display code
relative to the other element arrangements.

4AFC Task

Figures 24 and 25 depict the observers' accuracy (d')
and reaction time measures for the six display arrangements
in the 4AFC task. Again, three panels, a-c, in each figure
represent the individual subjects' data and the fourth
panel, d, is the average for the three observers. Each of
the subjects d' and reaction time measures are based on
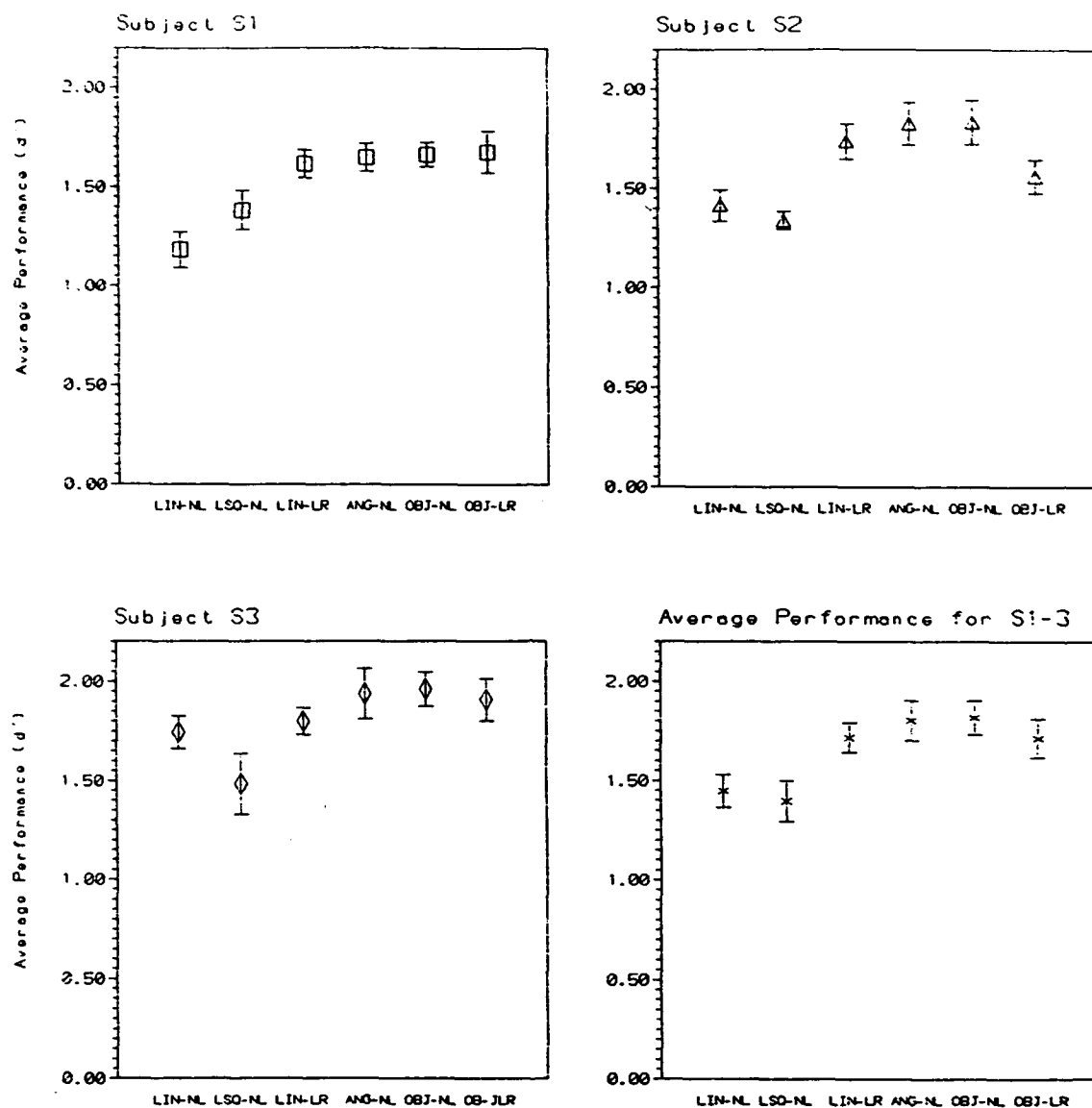eight blocks of 100 trials, and the error bars are one
standard error of the mean.
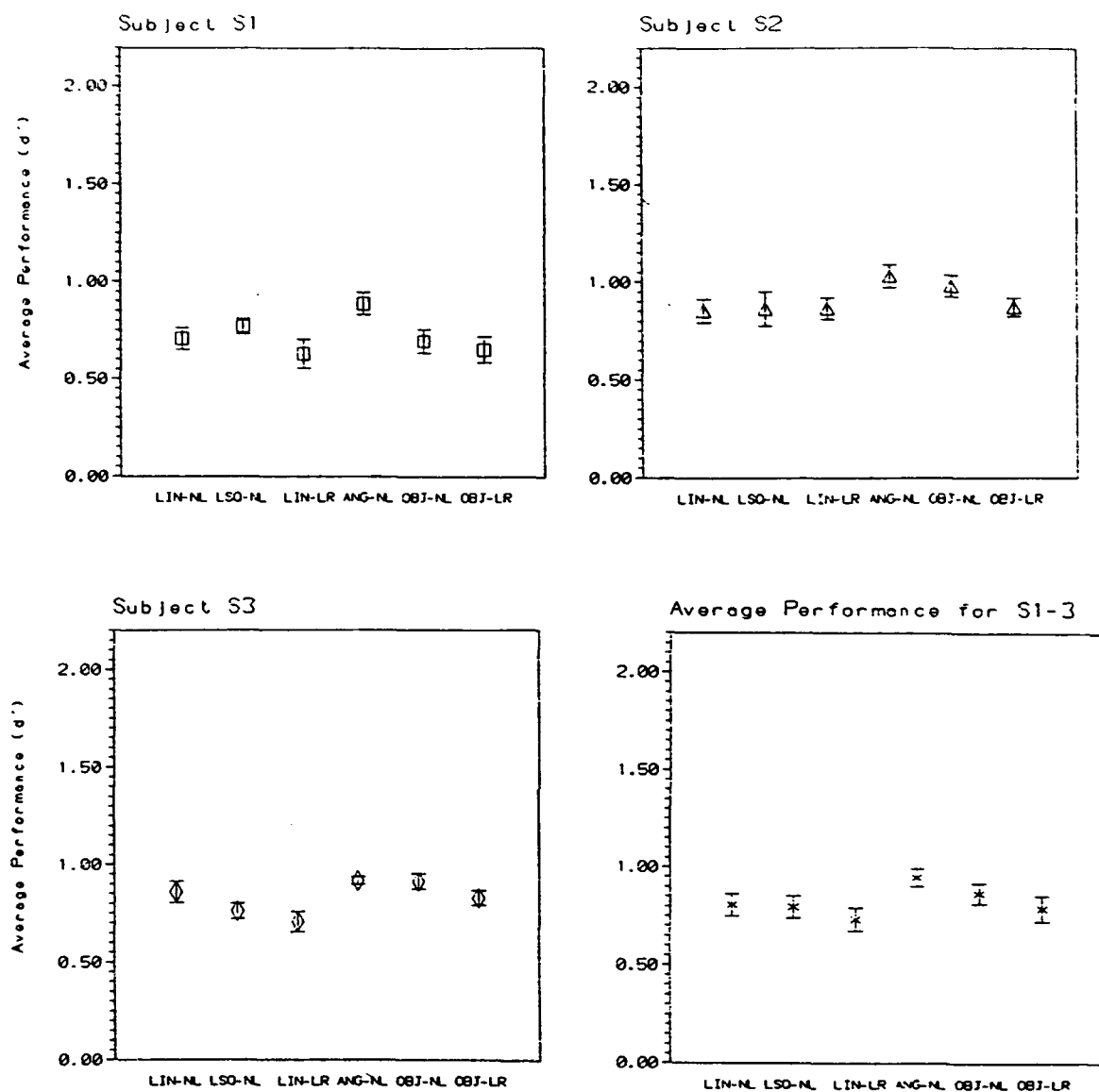
Figure 24. The observers' average performance (d') measures for the six arrangements in the 4AFC task. Panels a-c represent the individual subjects and panel d is the average data. The error bars are one standard error of the mean.

Figure 25. The observers' average reaction times measures
(measured from the offset of the mask to response) for
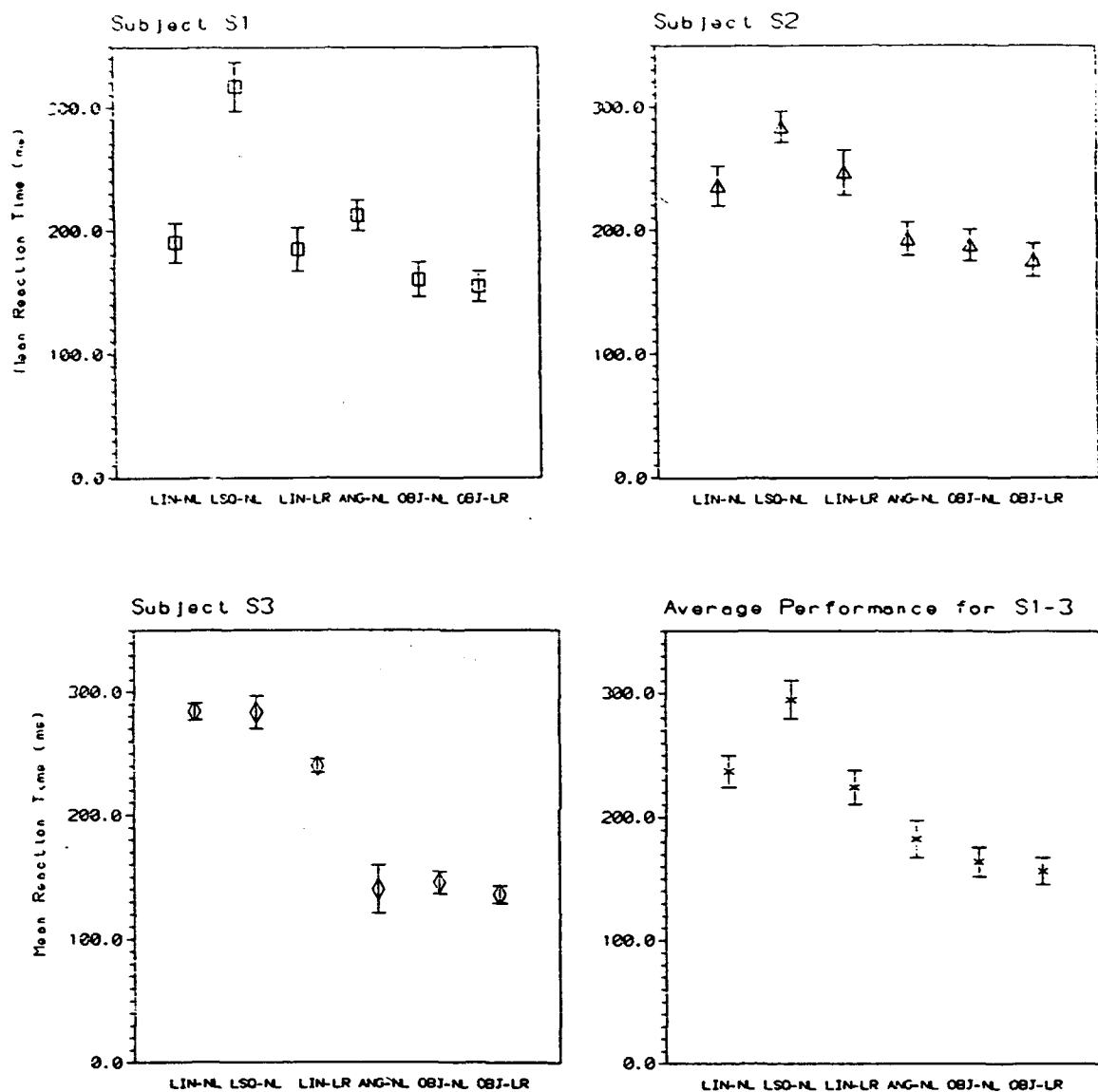the six conditions in the 4AFC task. Panels a-c repre-
sent the individual subjects and panel d is the average
data. The error bars are one standard error of the mean.

Again, separate repeated measures ANOVAs were performed on the observers' d' and reaction time measures, collapsed across the eight trial blocks. There was a significant effect of display arrangement ($F(5,10)=4.134, p \leq 0.05$) on observers' accuracy. Subsequent analytic comparisons, using the pooled variance for the error term, indicated that there was not an overall difference between displays with and without emergent features. However, there was a performance advantage ($F(1,10)=5.63, p \leq 0.05$) for display arrangements that produced an emergent feature with an enclosed contour (e.g., OBJ displays) relative to an emergent feature without this property (LIN-LR). There was also a significant ($F(1,10)=11.75, p \leq 0.01$) difference between the ANG-NL and the two line-graph display arrangements that did not possess an emergent feature (LIN-NL and LSQ-NL), though. Given this latter difference, these effects may be best characterized in terms of differences between line-graph and angular element coding.

Based on the accuracy data alone, observers tended to show poorer performance when emergent features were present. The observers showed the lowest performance for the LIN-LR display arrangement and highest performance for the ANG-NL display arrangement. In addition, among the angular display codes all observers showed lowest performance for the OBJ-LR arrangement.

However, the reaction time data did not completely support this pattern of effects. Unlike the accuracy data,

analysis of reaction times indicated that reaction time was faster ($\underline{F}(1,10)=11.6$, $\underline{p} \leq 0.01$) for display element arrangements that produced emergent features than those that did not possess these features. There was evidence for speed-accuracy tradeoffs among the separate linear and angular displays. For instance, among the linear displays the LIN-LR arrangement showed less accuracy, but faster reaction times than the other two line graph displays. However, there remained a performance advantage for angular element coding compared to the LIN-LR display arrangement.

Discussion

There is a great deal of interest from both theoretical and practical perspectives in how human detection performance is affected by element configuration. One of the main issues concerns whether emergent features produced by selected element arrangements help or hinder processing of the underlying elements. Evidence, so far, suggests that the impact on performance may depend upon which feature is most salient (Pomerantz, 1981) and how well this feature relates to the task being performed (Buttigieg & Sanderson, 1991).

This experiment compared performance in Yes/No and 4AFC detection tasks, for different arrangements of the line element components of simple visual displays. From the Yes/No task data, it was found that observer accuracy was affected by the display arrangement when observers were presented a line-graph display code. Observer accuracy was highest when line-graph display arrangements produced an

emergent feature (LIN-LR), and performance in this condition was indistinguishable from the angular display code arrangements which consistently produced superior performance.

Alternatively, the same feature appeared to hinder performance in the 4AFC task, which required focused attention to the separate elements. That is, accuracy was relatively low and reaction time was higher for the line-graph display arrangement which possessed an emergent feature. Similarly, there was a tendency for poorer performance given an emergent angular display, OBJ-LR, relative to an nonemergent angular display, ANG-NL. This pattern of effects would be expected if attention is automatically directed toward the emergent feature, and additional processing capacity has to be invoked to gather information from the underlying elements (Navon, 1977; Pomerantz & Schwaitzberg, 1975).

Although there is no strong evidence suggesting an effect of the relationship between the emergent feature and the decision statistic, it is not possible to completely rule out this factor. For instance, in the Yes/No task the emergent feature advantage observed for the LIN-LR arrangement relative to the other line graph displays could also be explained as a difference due to the relationship between the emergent feature and the decision statistic. This is true because the LIN-LR arrangement possessed both factors, where the other two displays possessed neither of these factors. It was only by comparing performance across tasks

that we could draw conclusions about which factors were influencing performance. Furthermore, some caution should be exercised when interpreting the results since the poor performance observed in the LIN-LR display in the 4AFC task may be related to masking effects of processing nearby items in the visual field (Eriksen & Eriksen, 1974).

Wickens and others have argued that this pattern of effects may be explained in terms of the proximity compatibility principle (Andre & Wickens, 1988; Carswell & Wickens, 1987; Wickens, 1992). According to the proximity compatibility principle, tasks that require integration are better supported by display arrangements which have high perceptual proximity; whereas, tasks which require focused attention are better supported by arrangements which have low perceptual proximity. If there were more distinguishable differences among the angular element display arrangements which produced consistently high performance, it may be possible to further eliminate alternative hypotheses.

This raises another question. Why was performance so good for displays composed of angular elements? One explanation for this angular display code advantage is that it is easier to extract magnitude information when it is coded as changes in the size of an angle than when it is coded as the length of a line segment. That is, the angle may emphasize the scale of the underlying element magnitude. Alternatively, observers may have used the direction of the angle rather than the size of the angle or the area enclosed by

the angle to make their decisions. For instance, observers may have identified angles pointing toward the center of the display (OBJ-LR) or downward (OBJ-NL or ANG-NL) as representing "noise," and the opposite as indicating "signal" events. Thus, their decisions would have been based on binary information rather than the actual magnitudes of the underlying elements.

A simple test of the latter possibility would be to conduct a study in which the tasks and displays are identical to those used in the preceding study. However, the distribution parameters would be manipulated so that in one case observers could use the direction of the angle and in the other cases they could not. For instance, if the underlying signal and noise distributions had relatively small or large means then most of the angles would be either small or large, respectively. Thus, the direction of the angles would be less useful in forming a detection decision. If observers are using the magnitude of the angles, there should be no differences among the selected pairs of means, as long as the distance between the distribution means and the standard deviations were held constant.

GENERAL CONCLUSIONS

We investigated whether selected display coding factors would assist observers in visual signal detection. The factors investigated were partially selected based on knowledge of the predicted optimal observer as defined in the TSD paradigm (Green & Swets, 1966). The evidence from these studies suggests potential means for coding displays that will assist observers in forming detection decisions.

In the first study, observers performed a Yes/No detection task where the separate sources differed in reliability. The main concern was to determine whether observers included these differences in source reliability in their detection decisions. Evidence from research on human decision making suggests that when informational sources differ in informativeness, decision makers generally do not consider these differences in forming their decisions. Instead, they acts as though the sources are equally informative (Wickens, 1984), and weight them accordingly. Berg (1990), also, found that observers were better at weighting sources equal rather than unequal in reliability in an auditory frequency discrimination task.

The results from the first study were consistent with prior evidence; observers were generally better at weighting sources equal rather than unequal in reliability. However,

when sources high in reliability were cued by gauge luminance, weighting efficiency was equivalent to the equal reliability condition. Furthermore, the best performance and highest weighting efficiency occurred when sources high in reliability were cued by gauge luminance, presented at long stimulus durations, and contiguous rather than distributed throughout the visual field. The performance advantage associated with grouped source reliabilities is consistent with the results of Posner et al. (1980).

Examination of the observers' weights indicate that observers tended to use a relatively equal weighting strategy when there was uncertainty about the location of the more reliable sources. This may partly explain the performance advantage for the odd arrangement in the non-cued conditions, since this arrangement was most similar to an equal reliability pattern. However, based on the current evidence it is not clear what factor accounts for this advantage, and why it is not maintained in the cued condition. A future study which examines other factors that may contribute to this advantage may help to understand these effects. For instance, is it the size of the difference between sources high and low in reliability or the distribution of sources reliabilities which produce this advantage?

Despite the relatively equal weighting pattern used in the non-cued conditions, observers showed moderate sensitivity to differences in source reliability. This was true even when observers could not use the trial-to-trial source

variability to identify which were more reliable (e.g. the UNcM condition). This may be related to the consistency in the vertical displacements of the gauge markers for the high reliability sources. The fairly straight line produced by the gauge markers common vertical positions in the visual field may have engaged observers attention, helping them to identify which sources were more reliable. Thus, an "emergent" pattern produced by the gauge marker arrangement may have assisted observers in this task.

Whether or not such "emergent" features assist observers in making detection decisions was addressed in the second study. The second study examined the effects of two factors related to display element arrangement on observers' detection decisions for both Yes/No and 4AFC detection tasks. The first factor was whether or not the display element arrangement produced an emergent feature. That is, a feature which is produced by the configuration of the underlying elements, but not present in any given element (Treisman, 1986). The second factor was whether or not the emergent feature had some property (e.g. size or area) that was directly related to the optimal decision statistic for the Yes/No detection task. Based on current evidence, many argue that the important factor is the relationship between the task and the display code (Bennett & Flach, 1992; Bennet, Toms, & Woods, 1993; Schmidt & Elvers, 1992; Sanderson et al., 1989).

Arranging the line graph displays to produce an emergent feature, improved Yes/No performance and impaired 4AFC performance. However, it was not possible to completely rule out the effects of the coded relationship between the optimal detection statistic and some property of the emergent feature. The angular display arrangements consistently produced high performance, and there were no distinguishable differences among the separate angular arrangements, for either task.

This angular advantage may be a function of at least two possible factors. First of all, it may be easier to extract magnitude information when it is coded as changes in the size of an angle than when it is coded as the length of a line segment. That is, the angle may emphasize the scale of the underlying element magnitude. Alternatively, the observers may have been able to use the direction of the angle to make their decisions about the underlying state of the system. A future study may provide some insights into whether or not these factors where, in fact, producing this performance advantage. Bennett, Toms, and Woods (1993) point out that emphasizing the scale of the underlying element magnitude helps observers to process the underlying elements. This is especially important when attention has to be focused on elements which are arranged to produce an emergent feature.

Thus, by defining the optimal observer we can (1) identify how well humans perform relative to the theoretical

ideal, and (2) identify means of aiding performance based on what we discover is causing inferior performance. For instance, introducing a luminance cue helps observers to prioritize sources according to their informativeness. Furthermore, using an angular display code in visual signal detection tasks can produce nearly ideal performance in both Yes/No and 4AFC detection tasks. Otherwise, designers should attempt to create display codes which possess "well-mapped" emergent features.

REFERENCES

Andre, A. D., & Wickens, C. D. (1988). The interaction of spatial and color proximity in aircraft stability information displays. In <u>Proceedings of the Human Factors Society 32nd Annual Meeting</u> (pp. 1371-1375). Santa Monica, CA: Human Factors Society.

Ankrum, C., & Palmer, J. (1991). Memory for objects and parts. <u>Perception and Psychophysics</u>, <u>50</u>, 141-156.

Bennett,K.B., and Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. <u>Human Factors</u>, 34(5), 513-533.

Bennett, K.B., Toms, M.L., & Woods, D.D. (1993). Emergent features and graphical elements: Designing more effective configural displays. <u>The Journal of the Human Factors Society</u>, <u>35</u>, 71-97.

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. <u>Journal of the Acoustical Society of America</u>, <u>86</u>,1743-1746.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. <u>Journal of the Acoustical Society of America</u>, <u>88</u>, 149-158.

Berg, B. G., & Green, D. M. (1990).Spectral weights in profile listening. <u>Journal of the Acoustical Society of America</u>, <u>88</u>, 758-766.

Boles, D. B., & Wickens, C. D. (1987). Display formatting in information integration and nonintegration tasks. <u>Human Factors</u>, <u>29</u>, 395-406.

Buttigieg, M. A., & Sanderson, P. M. (1991). Emergent features in visual display design for two type of failure detection tasks. <u>Human Factors</u>, 33(6), 631-651.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. <u>Ergonomics</u>, <u>30</u>, 511-527.

Christ, R. E. (1990). Review and analysis of color ʳoding research for visual displays. In M. Venturino (ed.) Selected readings in human factors (pp. 89 - 117). Santa Monica, CA: Human Factors Society.

Durlach, N.I., Braida, L.D., & Ito, Y. (1986). Toward a model for discrimination of broadband signals. Journal of the Acoustical Society of America, 80, 63-72.

Eriksen, B. A. & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. Perception and Psychophysics, 16, 143-149.

Eriksen, C. W., & Rohrbaugh J. (1970). Visual masking in multielement displays. Journal of Experimental Psychology: General, 83, 147-154.

Fisher, D. L., & Tan, K. C. (1989). Visual displays: The highlighting paradox. Human Factors, 31, 17-30.

Green, D.M. (1992). Signal detection theory. Unpublished document. Chapters 2 & 6.

Green, D. M., & Swets, J. A. (1966). Signal Detection Theory. New York: Wiley.

Hacker, M.J., & Radcliff, R. (1979). A revised table of d' for M-alternative forced choice. Perception & Psychophysics, 26, 168-170.

Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. Perception and Psychophysics, 46, 365-374.

MacMillian, N. A. & Creelman, C. D. (1991). Detection theory: A user's guide. Cambridge: Cambridge University Press.

Mitchell, J. A., & Biers, D. W. (1992). Decision statistic mapping and number of informational dimensions on decision making with graphical elements. Proceedings of the Human Factors Society 36th annual meeting (pp.1503 - 1507). Santa Monica, CA: Human Factors Society.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 9, 353-383.

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors, 33, 389-400.

Pomerantz, J. R. (1981). Perceptual Organization in Informa-
tion Processing. In M. Kubovoy & J. R. Pomerantz
(Eds.), Perceptual organization. Hillsdale, NJ: Erlbaum.

Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by
proximity: Selective attention measures. Perception and
Psychophysics, 18, 355-361.

Posner, M.I., Snyder, C. R. R., & Davidson, B. J. (1980).
Attention and the Detection of Signals. Journal of
Experimental Psychology: General, 109, 160-174.

Purcell, D. G. & Stewart, A. L. (1991). The object-detection
effect: Configuration enhances perception. Perception
and Psychophysics, 50, 215-224.

Reicher, G. M. (1969). Perceptual recognition as a function
of meaningfulness of stimulus materials. Journal of
Experimental Psychology, 81, 275-280.

Sanderson, P. M., Flach, J. M., Buttigieg, M. S., & Casey,
E. J. (1989). Object displays do not always support
better integration task performance. Human Factors, 31,
183-187.

Sanderson, P. M., Haskell, & Flack, J. M. (1992). The com-
plex role of perceptual organization in visual display
design theory. Ergonomics: Cognitive ergonomics III,
35, 1199-1219.

Schmidt, S., & Elvers, G. C, (1992). The influence of
color, closure, and correlation on integral and separa-
ble performance with object displays. Proceedings of
the Human Factors Society 36th Annual Meeting
(pp. 1488-1492). Santa Monica, CA: Human Factors
Society.

Schum, D.A. (1975). The weighting of testimony in judicial
proceedings from sources having reduced credibility.
Human Factors, 17, 172-182.

Sorkin, R. D., Mabry, T. R., Weldon, M. S, & Elvers, G. C.
(1991). Integration of information from multiple
element displays. Organizational Behavior and Human
Decision Processes, 49, 167-187.

Sorkin, R. D., Robinson, D. E., & Berg, B. G. (1987). A
detection theory method for the analysis of visual and
auditory displays. Proceedings of the Human Factors
Society 31st annual meeting (pp.1184-1187).
Santa Monica, CA: Human Factors Society.

Tanner, W. P., & Birdsall, T. G. (1958). Definitions of d'
and η as psychophysical measures, Journal of the
Acoustical Society of America, 30, 922-928.

Treisman, A. M. (1986). Properties, parts, and objects. In
K. Boff, L. Kaufmann, & J. Thomas (Eds.). Handbook of
perception and human performance (pp. 35-1-35-70).
New York: Wiley.

Weisstein, N., & Harris, C. S. (1974). Visual detection of
line segments: An object-superiority effect. Science,
186, 752-754.

Wickens, C.D. (1984). Engineering psychology and human
performance. Columbus, OH: Charles E. Merrill
Publishing Co.

Wickens, C. D., & Andre, A. D. (1990). Proximity compati-
bility and information display: Effects of color, space,
and objectness on information integration. Human
Factors, 32, 61-77.

Williams, L. J. (1982). Cognitive load and the functional
field of view. Human Factors, 24, 683-692.

Woods, D. D., Wise, J., & Hanes, L. (1981). An evaluation
of nuclear power plant safety parameter display systems.
Proceedings of the Human Factors Society 25th annual
meeting (pp.110-114). Santa Monica, CA: Human Factors
Society.

APPENDIX A
YES/NO DECISION STATISTIC

In a Yes/No detection task an observer is presented n independent sources of information, x_1, x_2, ..., x_n. On a given trial, each element represents one of two alternatives (signal or noise). On signal trials, all x_i are sampled from a distribution with the parameters μ_{si} and σ_{si}. Noise trials are sampled from a distribution with parameters μ_{ni} and σ_{ni}, where $\sigma_{si} = \sigma_{ni} = \sigma_{ei}$, and $\mu_{ni} < \mu_{ni}$.

The likelihood that this evidence represents a signal is related to the probability of their joint occurrence,

$$L(x_1, x_2, \ldots, x_n) = L(x_1) * L(x_2) \ldots * L(x_3). \tag{A1}$$

The likelihood ratio for a given source, x_i, is the ratio of the conditional probabilities for that source. That is,

$$L(x_i) = \frac{f(x_i/s)}{f(x_i/n)} = \frac{[1/(2\pi\sigma_{ei}^2)^{\frac{1}{2}}]EXP[-\frac{1}{2}((x_i-\mu_{si})/\sigma_{ei})^2]}{[1/(2\pi\sigma_{ei}^2)^{\frac{1}{2}}]EXP[-\frac{1}{2}((x_i-\mu_{ni})/\sigma_{ei})^2].} \tag{A2}$$

Once the equation is reduced, based on the general laws of exponents we have the following:

$$L(x_i) = EXP[-\frac{1}{2}(((x_i-\mu_{si})/\sigma_{ei})^2 - ((x_i-\mu_{ni})/\sigma_{ei})^2)]. \tag{A3}$$

By taking the natural logarithm of the likelihood ratio, and, again reducing the equation we have the following:

$$lnL(x_i) = x_i((\mu_{si}-\mu_{ni})/\sigma_{ei}^2) - \frac{1}{2}((\mu_{si}^2-\mu_{ni}^2)/\sigma_{ei}^2). \tag{A4}$$

Then, for the combined evidence, $l(x_1, x_2, x_3)$ it turns out that a weighted sum of the evidence, x_i, is directly related to the likelihood ratio,

$$lnL(x_1, x_2, \ldots, x_n) = \sum_{1}^{n}((\mu_{si}-\mu_{ni})/\sigma_{ei}^2)x_i - \frac{1}{2}((\mu_{si}^2-\mu_{ni}^2)/\sigma_{ei}^2). \tag{A5}$$

APPENDIX B
BERG'S THEORETICAL SOLUTION OF THE WEIGHTS

Appendix A showed that the optimal decision statistic for a Yes/No task is a weighted sum of the evidence (Green & Swets, 1966). Berg's Conditional-On-A-Single-Stimulus (COSS) technique is based on this assumption. That is, the observer responds "signal" when the evidence, x_i, weighted by some arbitrary number, a_i, surpasses some criterion value, D,

$$\sum_{1}^{n} a_i x_i > D, \tag{B1}$$

otherwise the observer responds "noise." In the theoretical solution of the individual weights Berg begins by isolating a single element, x_i, on the left side of equation B1, producing the following inequality:

$$x_i > D - [\sum_{j}^{n} a_j x_j / a_i], \qquad \text{given } i \neq j. \tag{B2}$$

Then, a new variable, Y_i, is substituted for the right side of the inequality. Y_i is also normally distributed since it represents the sum of independent, normally distributed, random variables, x_i, and has the parameters:

$$E[Y_i] = D - [\sum_{j}^{n} a_j x_j / a_i]$$

$$\tag{B3}$$

$$Var[Y_i] = \sum_{j}^{n} a_j^2 \sigma_{ej}^2 / a_i^2.$$

Given a signal trial and a particular source, x_i, the probability of saying "signal" given x_i is

$$P(S|x_i,s) = \int_{-\infty}^{x} f(u|s)\,du, \qquad (B4)$$

where $f(x|s)$ is a normal density function with a mean = $E[Y_i|s]$ and a variance = $VAR[Y_i|s]$. A similar density function exits for noise trials.

Berg uses the sum of the variance of the normal density, $VAR[Y_i]$, and the variance of the original x_i, σ_{ei}^2, to derive the estimated weights. One element, in this case x_j, is set to unity to solve for the remaining weights (k=1..m) as shown below,

$$\frac{VAR[Y_j] + \sigma_{ej}^2}{VAR[Y_k] + \sigma_{ek}^2} = \frac{\sum_1^n a_j^2 \sigma_{ej}^2 / a_i^2}{\sum_1^n a_i^2 \sigma_{ei}^2 / a_k^2} = \frac{a_k^2}{a_j^2}. \qquad (B5)$$

In the actual derivation of the weights, we know the values of σ_{ei}^2. We need to find the values of $VAR[Y_i]$ given the two types of trials (signal and noise), and we need to specify which element is set to unity. The values $VAR[Y_i]$ are estimated by finding the variances of the cumulative normals with the best Chi-square fits to the observer's COSS functions. Selection of the element to be set to unity is arbitrary; the center element, x_5, was selected in this study.

APPENDIX C
4AFC DECISION STATISTIC

In a Four-Alternative-Forced-Choice task an observer is presented four independent sources of information, x_1, x_2, x_3, x_4.   On a given trial, three sources represent noise and their values are drawn from a normal distribution with a mean, μ_n, and a standard deviation, σ_n. The value of the remaining source is drawn from a distribution with a mean, μ_s and a standard deviation, σ_s, where $\mu_n < \mu_s$, and $\sigma_s = \sigma_n = \sigma_e$. The observer has to decide which source represents the signal event. That is, in a 4AFC task, the observer has to decide which of the following alternatives is true: $<s,n,n,n>$, $<n,s,n,n>$, $<n,n,s,n>$, or $<n,n,n,s>$.

To simplify the derivation of the optimal decision statistic, Green (1992) considers the four alternatives as representing four possible signals (for instance $<s,n,n,n>$ would equal Sg_1) which are compared with noise alone, $<n,n,n,n>$. That is, if $x = <x_1, x_2, x_3, x_4>$ then

$$l(x|Sg_i) = \frac{f(x_1/s)\,f(x_2/n)\,f(x_3/n)\,f(x_4/n)}{f(x_1/n)\,f(x_2/n)\,f(x_3/n)\,f(x_4/n).} \qquad (C1)$$

Once we substitute the definition for the conditional probabilities and generalize the equation for all possible Sg_i, equation C1 becomes

$$l(x|Sg_i) = \frac{M\,\prod_{j}^{m} EXP[-\tfrac{1}{2}((x_j-\mu_n)/\sigma_e)^2]\,EXP[-\tfrac{1}{2}((x_i-\mu_s)/\sigma_e)^2]}{M\,\prod_{j}^{m} EXP[-\tfrac{1}{2}((x_j-\mu_n)/\sigma_e)^2]\,EXP[-\tfrac{1}{2}((x_i-\mu_n)/\sigma_e)^2],} \qquad (C2)$$

101

where $j \neq i$, $M = (1/(2\pi\sigma_e^2)^{m/2})$, and $m = 4$. Once we reduce the equation we have the following:

$$l(x|Sg_i) = EXP[x_i((\mu_s-\mu_n)/\sigma_e^2) - \tfrac{1}{2}((\mu_s^2 - \mu_n^2 /\sigma_e^2))]. \qquad (C3)$$

Thus, the optimal decision statistic is to choose the largest value, x_i, since it is directly related to the largest likelihood ratio.

APPENDIX D
RELATIONSHIP BETWEEN YES/NO d' AND mAFC PERCENT CORRECT

Green (1992) shows that the integral in equation 19 can be expressed in terms of a Yes/No ROC since the same probability density functions, $f(x|s)$ and $f(x|n)$, can be used to define the hit and false alarm probabilities. If we let a specific signal value, x_s, equal u the false alarm probability is equal to

$$P_u(S|n) = \int_u^\infty f(x|n)\,dx, \qquad (D1)$$

and the hit probability is equal to

$$P_u(S|s) = \int_u^\infty f(x|s)\,dx. \qquad (D2)$$

The complements of these values are

$$1 - P_u(S|n) = \int_{-\infty}^u f(x|n)\,dx, \text{ and} \qquad (D3)$$

$$1 - P_u(S|s) = \int_{-\infty}^u f(x|s)\,dx, \text{ respectively.} \qquad (D4)$$

Taking the derivative of equation D4, we have the following,

$$-dP_u(S|s) = f(u|s)\,du \qquad (D5)$$

Then by substituting equations D3 and D5 into equation 19. Green (1992) produces the following equation:

$$P_2(C) = \int_{-\infty}^\infty [1-P_u(S|n)]\,-dP_u(S|s). \qquad (D6)$$

When the criterion value is low, u is $-\infty$, then the hit probability will be high, $P(S/s) = 1$. Similarly, when the criterion value is high, $u = \infty$, then the hit probability is

low, $P(S/s) = 0$. So the limits of integration can be replaced by these values, and their order is switched by changing the sign of the integral, leading to the following equation:

$$P_2(C) = \int_0^1 [1-P_u(S|n)]dP_u(S|s) \qquad (D7)$$

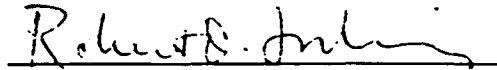where the right side of this equation is equal to the area under a Yes/No ROC.

Green (1992), then shows that equation D7 can be rewritten for m alternatives, where $m > 2$. For an m-alternative forced-choice task, as the one used in the second study, there are m-1 noise samples. The probability of a correct response in this case is equal to the probability that the signal sample, x_s (or in this discussion u), is greater than all m-1 noise samples. Thus, equation D7 is rewritten to account for the m-1 noise samples giving us the following equation:

$$P_m(c) = \int_0^1 [1-P_u(S|n)]^{m-1}dP_u(S|s). \qquad (D8)$$
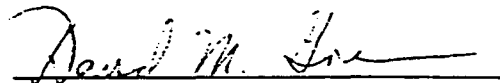
BIOGRAPHICAL SKETCH

DeMaris Montgomery was born May 7, 1965, in Fremont Ohio. She received her B.S. degree in psychology 1987 from the University of Dayton, and her M.S. (1991) and Ph.D. degrees from the University of Florida. DeMaris' research interests are best characterized as the study of human visual performance, perception, and psychophysics.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
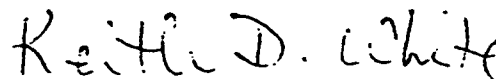
Robert D. Sorkin, Chairman
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
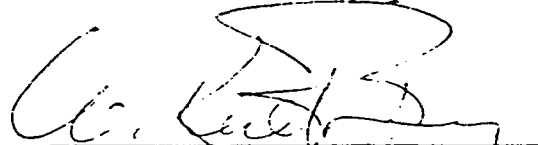
David M. Green
Graduate Research
    Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
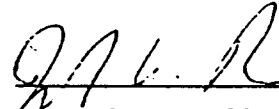
Keith D. White
Associate Professor of
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

W. Keith Berg
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph W. Alba
Associate Professor of
Marketing

This dissertation was submitted to the Graduate Faculty of the Department of Psychology in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1993   .

_____
Dean, Graduate School